

Comparison of Sequential data assimilation methods for the Kuramoto-Sivashinsky equation

M. Jarda , I. M. Navon & M. Zupanski

*Department of Scientific Computing
Florida State University
Tallahassee, FL 32306-4120, USA*

and

*Cooperative Institute for Research in the Atmosphere,
Colorado State University, 1375 Campus Deliver,
Fort Collins, CO 80523-1375, USA*

SUMMARY

The Kuramoto-Sivashinsky equation plays an important role as a low-dimensional prototype for complicated fluid dynamics systems having been studied due to its chaotic pattern forming behavior. Up to now, efforts to carry out data assimilation with this 1-d model were restricted to variational adjoint methods domain and only Chorin and Krause [26] tested it using a sequential Bayesian filter approach. In this work we compare three sequential data assimilation methods namely the Kalman filter (EnKF) approach the sequential Monte-Carlo particle filter approach (PF) and the Maximum Likelihood Ensemble Filter methods (MLEF). This comparison is to the best of our knowledge novel. We compare in detail their relative performance for both linear and nonlinear observation operators. The results of these sequential data assimilation tests are discussed and conclusions are drawn as to the suitability of these data assimilation methods in the presence of linear and nonlinear observation operators. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: *Sequential data assimilation, ensemble Kalman filter, particle filter, Kuramoto-Sivashinsky equation.*

1. Introduction

Sequential data assimilation is a relatively novel and versatile multidisciplinary methodology. It combines observations of the current (and possibly, past) state of a system with results from a mathematical model (the forecast) to produce an analysis, providing "the best" estimate of the current state of the system. Central to the concept of sequential estimation data assimilation is the propagation of flow dependent covariance errors.

In sequential estimation, the analysis and forecasts can be viewed of as probability distributions. The analysis step is an application of the Bayes theorem. Advancing the probability distribution in time, in the general case is done by the Chapman-Kolmogorov equation, but since it is unrealistically expensive, various approximations operating on representations of the probability distributions are used instead. If the probability distributions

are normal, they can be represented by their mean and covariance, which gives rise to the Kalman filter (KF). However it is not computationally feasible to store the covariance due to the large number of degrees of freedom in the state, so various approximations based on Monte-Carlo ensemble calculations are used instead.

In the present paper we compare standard ensemble Kalman filter (EnKF) along with the particle filter (PF) using various resampling strategies to the Maximum Likelihood Ensemble Filter(MLEF) of Zupanski [66]. In addition, we compare the performance of each of the above mentioned methods in the presence of linear and nonlinear observation operators.

While work on 4-D variational data assimilation using as model the Kuramoto-Sivashinsky equation has been carried out by very few research workers (see Protas et al. [19]), to the best of our knowledge only the work of Chorin and Krause [26] has used Bayesian filters.

We further discuss issues of computational efficiency for the above mentioned filters in terms of CPU time and memory requirements for the three aforementioned methods taking into account the additional effort required by PF due to various resampling strategies aimed at preventing filter degeneracy.

The structure of the present work is as follows. After the introduction, we present in section 2 the K-S equation and its numerical solution. In section 3 we present the formulation of the PF and issues related to various resampling strategies implemented. The formulation of the standard EnKF is presented as well. Then we present the formulation of the MLEF method viewed as a hybrid between sequential and 4-D variational data assimilation. Section 4 provides discussion of the numerical experiments followed by a discussion of their significance. Finally, section 5 contains summary and conclusions along with directions for further research.

2. The Kuramoto-Sivashinsky Equation

The Kuramoto-Sivashinsky (K-S) equation is an evolution equation in one space dimension, with a Burgers nonlinearity, a fourth order dissipation term and a second anti-dissipative term. It assumes the form

$$u_t + u_{xxxx} + u_{xx} + uu_x = 0, \quad (x, t) \in \mathbb{R} \times \mathbb{R}^+ \quad u(x, 0) = u_0(x), x \in \mathbb{R} \quad (1)$$

The K-S equation models pattern formations in different physical contexts and is a paradigm of low-dimensional behavior in solutions to partial differential equations. It arises as a model amplitude equation for inter-facial instabilities in many physical contexts. It was originally derived by Kuramoto and Tsuzuki ([6], [7]) to model small thermal diffusive instabilities in laminar flame fronts in two space dimensions. It has also been derived in the context of angular-phase turbulence for a system of reaction-diffusion modeling the Belousov-Zhabotinskii reaction in three space dimensions. Sivashinsky([8],[9]) derived it independently to model small thermal diffusive instabilities in laminar flame fronts. The equation also arises in modeling small perturbations from a reference Poiseuille flow of a film layer on an inclined plane [10], while Babchin et al.[11] derived (1) as a general mechanism modeling the nonlinear saturation of instabilities in flow films as in the Rayleigh-Taylor-type instability.

The K-S equation is non-integrable, and no explicit solutions exist. It is characterized by a second-order unstable diffusion term, responsible for an instability at large scales, a fourth-order stabilizing viscosity term, which provides damping at small scales; and a quadratic nonlinear

coupling term which stabilizes by transferring energy between large and small scales. This is readily apparent in Fourier space, where one may write(1)with periodic boundary condition as

$$\frac{d\hat{u}_k}{dt} = (k^2 - k^4)\hat{u}_k + \frac{i}{2} \sum_{k' \in \mathbb{Z}} k' \hat{u}_{k'} \hat{u}_{k-k'} \tag{2}$$

where

$$u(x, t) = \sum_{k \in \mathbb{Z}} \hat{u}_k(t) \exp(ikx), \quad k = n \frac{2\pi}{L}, \quad k' = m \frac{2\pi}{L}, \quad m, n \in \mathbb{Z}, \quad i = \sqrt{-1},$$

The zero solution is linearly unstable to modes with $|k| < 1$; whose number is proportional to the bifurcation parameter L , are coupled to each other and to damped modes at $|k| > 1$ through the non-linear term. In an effort to characterize, understand and predict the spatially and temporally nontrivial dynamical behavior including chaos of K-S, numerical simulation has been conducted by Hyman and Nicolaenko [12] and by Kevrekidis et al.[13] to cite but few.

Foias et al. [14], Nicolaenko et al.[15] and Temam [5] established the existence of a unique compact inertial manifold for the K-S equation. They also demonstrated that the K-S equation is strictly equivalent to a low-dimensional dynamical system. That is, all orbits are attracted exponentially to a finite-dimensional, bounded, compact, smooth manifold and the dynamics take a place in this “inertial” manifold. This implies that the transition to chaos of the K-S equation can be analyzed using the tools developed for low-dimensional dynamics systems.

Desertion and Kazantzis [17] used the K-S equation to examine the performance improvement of a class of nonlinear transport processes subject to spatio-temporally varying disturbances through the employment of a comprehensive and systematic actuator activation policy. Lee and Tran [18] obtained a reduced-order system that can accurately describe the dynamics of the K-S equation by employing an approximate inertial manifold and a proper orthogonal decomposition. From this resulting reduced-order system, they designed and synthesized the feedback controller for the K-S equation. Recently, Protas et al.[19] came up with a comprehensive framework for the regularization of adjoint analysis in multiscale PDE systems. They examined the regularization opportunities available in the adjoint analysis and optimization of multiscale, and applied the proposed regularization strategies to the K-S equation

2.1. Mathematical Formulation

We consider the solutions of

$$u_t + u_{xxxx} + u_{xx} + uu_x = 0, \quad (x, t) \in \mathbb{R} \times \mathbb{R}^+ \quad u(x, 0) = u_0(x), x \in \mathbb{R} \tag{3}$$

which are space periodic of period L , $u(x, t) = u(x + L, t)$, $L > 0$.

Let $\Omega \subset \mathbb{R}$, we denote by

$$\mathbb{L}^2(\Omega) = \left\{ u|u : \Omega \longrightarrow \mathbb{R}, u \text{ measurable and } \int_{\Omega} |u(x)|^2 dx < \infty \right\},$$

the Hilbert space of square integrable function over Ω endowed with the norm

$$\|u\|_2 = \left\{ \int_{\Omega} |u(x)|^2 dx \right\}^{1/2},$$

and by

$$\mathbb{H}^4(\Omega) = \left\{ u|u : \Omega \longrightarrow \mathbb{R}, u \text{ measurable and } \frac{d^i u}{dx^i} \in \mathbb{L}^2(\Omega), \text{ for } i = 0, \dots, 4 \right\},$$

the fourth-order Sobolev space with the norm, see Adams [1] and also Mazya [2],

$$\|u\|_{4,2} = \left\{ \|u\|_2^2 + \sum_{i=1}^4 \left\| \frac{d^i u}{dx^i} \right\|_2^2 \right\}^{1/2}.$$

Finally, we denote by $\mathbb{H}_{per}^4(\Omega)$ the closure of $\mathcal{C}_{per}^\infty(\Omega)$ for the \mathbb{H}^4 -norm. $\mathcal{C}_{per}^\infty(\Omega)$ is the subset of $\mathcal{C}^\infty(\mathbb{R})$ of Ω -periodic functions.

We set

$$\mathcal{A} = \frac{d^4}{dx^4}, \quad \mathcal{H} = \left\{ u \in \mathbb{L}^2\left(\frac{-L}{2}, \frac{L}{2}\right) \right\}$$

and

$$D(\mathcal{A}) = \mathbb{H}_{per}^4\left(\frac{-L}{2}, \frac{L}{2}\right) \cap \mathcal{H}, \quad \mathcal{V} = D(\mathcal{A}^{1/2}).$$

\mathcal{H} is a Hilbert space, the dissipative operator \mathcal{A} is a linear self-adjoint unbounded operator in \mathcal{H} with domain $D(\mathcal{A})$ and dense in \mathcal{H} . Assuming \mathcal{A} positive closed and that \mathcal{A}^{-1} is compact, \mathcal{V} is a Hilbert space endowed with the norm $|\mathcal{A}^{1/2} \cdot|$. Using Leray's method (see Lions[3], Temam[4]), the K-S equation (1) with initial condition $u_0 \in \mathcal{H}$ has a unique solution defined for all $t > 0$ and such that

$$u \in \mathcal{C}(\mathbb{R}^+; \mathcal{H}) \cap \mathbb{L}^2(0, T; \mathcal{V}) \quad \forall T > 0.$$

Moreover, if $u_0 \in \mathcal{V}$, then

$$u \in \mathcal{C}(\mathbb{R}^+; \mathcal{V}) \cap \mathbb{L}^2(0, T; D(\mathcal{A})) \quad \forall T > 0.$$

Furthermore, it has been proved in Nicoleanko et al. ([15],[16]) that only the odd solutions of (1) are stable for large t . Consequently, the subspace \mathcal{H} is restricted to

$$\mathcal{H} = \left\{ u \in \mathbb{L}^2\left(\frac{-L}{2}, \frac{L}{2}\right), u \text{ is odd} \right\}$$

3. Data Assimilation for the K-S Equation

Data assimilation is the process by which observational data distributed in space and time are fused with mathematical model forecast information aimed at obtaining the best initial conditions that are as near as possible to observations while satisfying model forecast as a strong constraint. The probabilistic state space formulation and the requirement for the updating of information when new observations are encountered are ideally suited for the Bayesian approach, and thus constitute an appropriate framework for data assimilation. The Bayesian approach and in particular ensemble or particle filtering methods are a set of efficient and flexible Monte-Carlo methods to solve the optimal filtering problem. Here one attempts to construct the posterior probability density function (PDF) of the state based on all available

information, including the set of received observations. Since this PDF embodies all available statistical information, it may be considered to be a complete solution to the estimation problem.

In the field of data assimilation, there are only few contributions in sequential estimation (EnKF, PF and MLEF) using the K-S equation. Chorin and Kruse [26] used particle filters and proposed a strategy for reducing the size of the system of equations to be solved for evolving the particles by adaptively finding subsets of variables that do not need to be recomputed and solving only in directions in which the dynamic system is expanding. In the work of Hu and Temam [27], a robust boundary control method for the K-S equation has been proposed, and a data assimilation problem corresponding to the K-S equation has been considered.

3.1. Sequential Bayesian Filter- theoretical setting

The sequential Bayesian filter employs a large number N of random samples or “particles” advanced in time by a stochastic evolution equation, to approximate the probability densities. In order to analyze and make inference about the dynamic system at least a model equation along with an observation operator are required. First, a model describing the evolution of the state with time, and an observation operator for noisy observations of the state. Generically, stochastic filtering problem is a dynamic system that assumes the form

$$\dot{\mathbf{x}}_t = f(t, \mathbf{x}_t, u_t, \mathbf{v}_t) \quad (4)$$

$$\mathbf{z}_t = h(t, \mathbf{x}_t, u_t, \mathbf{n}_t) \quad (5)$$

The equation (4) is the state equation or the system model, (5) is the observation operator equation, \mathbf{x}_t is the state vector, \mathbf{z}_t the observation vector and u_t is the system input vector serving as the driving force. \mathbf{v}_t and \mathbf{n}_t are the state and observation noises, respectively. In practical application, however, we are more concerned about the discrete-time filtering, and we consider the evolution of the state sequence $\{\mathbf{x}_k, k \in \mathbb{N}\}$, given by

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}), \quad (6)$$

where the deterministic mapping $f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \longrightarrow \mathbb{R}^{n_x}$ is a possibly non-linear function of the state \mathbf{x}_{k-1} , $\{\mathbf{v}_{k-1}, k \in \mathbb{N}\}$ is an independent identically distributed (i.i.d) process noise sequence, n_x, n_d are dimensions of the state and process noise vectors, respectively, and \mathbb{N} is the set of the natural numbers. The objective is to recursively estimate \mathbf{x}_k from observations

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{n}_k), \quad (7)$$

where $h_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_n} \longrightarrow \mathbb{R}^{n_z}$ is a possibly non-linear function, $\{\mathbf{n}_k, k \in \mathbb{N}\}$ is an i.i.d. observation noise sequence, and n_x, n_n are dimensions of the state and observation noise vectors, respectively.

We denote by $\mathbf{z}_{1:k}$ the set of all available observations \mathbf{z}_i up to time $t = k$, $\mathbf{z}_{1:k} = \{\mathbf{z}_i | i = 1, \dots, k\}$. From a Bayesian point of view, the problem is to recursively calculate some degree of belief in the state \mathbf{x}_k at time $t = k$, taking different values, given the data $\mathbf{z}_{1:k}$ up to the time $t = k$. Then the Bayesian solution would be to calculate the PDF $p(\mathbf{x}_k | \mathbf{z}_{1:k})$. This density will encapsulate all the information about the state vector \mathbf{x}_k that is contained in the observations $\mathbf{z}_{1:k}$ and the prior distribution for \mathbf{x}_k .

Suppose that the required PDF $p(\mathbf{x} | \mathbf{z}_{1:k-1})$ at time $k - 1$ is available. The prediction stage

uses the state equation (6) to obtain the prior PDF of the state at time k via the Chapman-Kolmogorov equation

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1}. \quad (8)$$

The probabilistic model of the state evolution, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, is defined by the state equation (6) and the known statistics of \mathbf{v}_{k-1} .

At time $t = k$, a measurement \mathbf{z}_k becomes available, and it may be used to update the prior via the Bayes rule

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}, \quad (9)$$

where the normalizing constant

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k. \quad (10)$$

depends on the likelihood function $p(\mathbf{z}_k|\mathbf{x}_k)$, defined by the measurement equation (7) and the known statistics of \mathbf{n}_k .

The relations (8) and (9) form the basis for the optimal Bayesian solution. This recursive propagation of the posterior density is only a conceptual solution. One cannot generally obtain an analytical solution. Solutions exist only in a very restrictive set of cases like that of the Kalman filters for instance. (namely, if f_k and h_k are linear and both v_k and n_k are Gaussian). Particle filters provide a direct approximation to the Bayes rule outlined above.

3.2. Particle Filters

Particle filters (see [30], [31], [32] and [28]) approximate the posterior densities by population of states. These states are called "particles". Each of the particles has an assigned weight, and the posterior distribution can then be approximated by a discrete distribution which has support on each of the particles. The probability assigned to each particle is proportional to its weight. The different (PF) algorithms differ in the way that the population of particles evolves and assimilates the incoming observations. A major drawback of particle filters is that they suffer from sample degeneracy after a few filtering steps. The PF suffers from "the curse of dimensionality" requiring computations that increase exponentially with dimension as pointed out by Silverman [60]. This argument was enhanced and amplified by the recent work of Bengtsson et al. [61] and Bickel et al. [62] and finally explicitly quantified by Snyder et al. [63]. They indicated that unless the ensemble size is greater than $\exp(\tau^2/2)$, where τ^2 is the variance of the observation log-likelihood, the PF update suffers from a "collapse" in which with high probability a single member is assigned a posterior weight close to one while all other members have vanishing small weights. This issue become more acute as we move to higher spatial dimensions.

Nevertheless in this 1-D problem the common remedy is to resample the prior PDF whenever the weights focus on few members of the ensemble assuming that our experiments do not encounter the exponential collapse. Here we use several strategies such as Systematic Resampling (SR), Residual Resampling (RR) and or the Bayesian bootstrap filter of Gordon et al. [33] see also Berliner and Wikel([28],[29]). Multinomial Resampling (MR) The SIR algorithm generates a population of equally weighted particles to approximate the posterior at some time

k . This population of particles is assumed to be an approximate sample from the true posterior at that time instant.

The PF algorithm proceeds as follows:

- **Initialization:** The filter is initialized by drawing a sample of size N from the prior at the initial time. The algorithm is then started with the filtering step.
- **Preliminaries:** Assume that $\{\mathbf{x}_{k-1}^i\}_{i=1,\dots,N}$ is a population of N particles, approximately distributed as in an independent sample from $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$
- **Prediction** Sample N values, $\{q_k^1, \dots, q_k^N\}$, from the distribution of \mathbf{v}_k . Use these to generate a new population of particles, $\{\mathbf{x}_{k|k-1}^1, \mathbf{x}_{k|k-1}^2, \dots, \mathbf{x}_{k|k-1}^N\}$ via the equation

$$\mathbf{x}_{k|k-1}^i = f_k(\mathbf{x}_{k-1}^i, \mathbf{v}_k^i) \quad (11)$$

- **Filtering:** Assign each $\mathbf{x}_{k|k-1}^i$, a weight q_k^i . This weight is calculated by

$$q_k^i = \frac{p(\mathbf{z}_k|\mathbf{x}_{k|k-1}^i)}{\sum_{j=1}^N p(\mathbf{z}_k|\mathbf{x}_{k|k-1}^j)} \quad (12)$$

This defines a discrete distribution which, for $i \in \{1, 2, \dots, N\}$, assigns probability mass q_k^i to element $\mathbf{x}_{k|k-1}^i$

- **Resampling:** Resample independently N times, with replacement, from the distribution obtained in the filtering stage. The resulting particles, $\{\mathbf{x}_k^i\}_{i=1,\dots,N}$, form an approximate sample from $p(\mathbf{x}_k|\mathbf{z}_{1:k})$.

The method outlined above can be justified as follows. If the particles at time $t = k - 1$ were an i.i.d sample from the posterior at time $t = k - 1$, then the predictive stage just produces an i.i.d. sample from the prior at time $t = k$. The filtering stage can be viewed as an importance sampling approach to generate an empirical distribution which approximates the posterior.

The proposal density is just the prior $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$, and as a result of Bayes formula, we obtain

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \mathbf{z}_k) \propto p(\mathbf{x}_k|\mathbf{z}_{1:k-1})p(\mathbf{z}_k|\mathbf{x}_k), \quad (13)$$

the weights are proportional to the likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$. As N tends to infinity, the discrete distribution which has probability mass q_i at point $\mathbf{x}_{k|k-1}^i$, converges weakly to the true posterior. The resampling step is a crucial and computationally expensive part in a particle filter. It is used to generate equally weighted particles aimed at avoiding the problem of degeneracy of the algorithm, that is, avoiding the situation that all but one of the weights are close to zero. The resampling step modifies the weighted approximate density $p(\mathbf{x}_k|\mathbf{z}_k)$ to an unweighted density $\hat{p}(\mathbf{x}_k|\mathbf{z}_k)$ by eliminating particles having low importance weights and by multiplying particles having highly importance weights. Formally:

$$p(\mathbf{x}_k|\mathbf{z}_k) = \sum_{i=1}^N q_i \delta(\mathbf{x}_k - \mathbf{x}_k^i) \quad (14)$$

is replaced by

$$\hat{p}(\mathbf{x}_k|\mathbf{z}_k) = \sum_{i=1}^N \frac{1}{N} \delta(\mathbf{x}_k - \mathbf{x}_k^*) = \sum_{i=1}^N \frac{n_i}{N} \delta(\mathbf{x}_k - \mathbf{x}_k^i) \quad (15)$$

where n_i is the number of copies of particle \mathbf{x}_k^i in the new set of particles $\{\mathbf{x}_k^*\}$.

Generically, it is implemented as follows:

- Draw N particles $\{\tilde{\mathbf{x}}_k^i\}_{i=1,\dots,N}$ from the uniform distribution.
- Assign the resampled particles $\{\tilde{\mathbf{x}}_k^i\}_{i=1,\dots,N}$ to $\{\mathbf{x}_k^i\}_{i=1,\dots,N}$ and assign equal weights $\frac{1}{N}$ to each particle.

3.3. Merging Particle Filter

To avoid the high cost inherent in recursive resampling (used in order to remove numbers with very small weights and replenish the ensemble), we test here a method put forward by Nakano et al. [57] called the Merging Particle Filter (MPF) which is a modification to the PF. The MPF is similar to the genetic algorithm of Goldberg [59]. A filtered ensemble is constructed based on samples from a forecast ensemble as in the regular PF with resampling. However, each particle of a filtered ensemble is generated as an amalgamation of multiple particles (i.e. ensembles of various sizes). The merging of several particles of a prior ensemble allows the degeneracy problem to be alleviated. For further details the reader is referred to the paper of Nakano et al. [57]

3.4. The Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) was first proposed by Evensen [37] and further developed by Burgers et al. [38] and Evensen ([39],[40]). It is related to particle filters in the context that a particle is identical to an ensemble member. EnKF is a sequential filter method, which means that the model is integrated forward in time and, whenever observations are available, these are used to reinitialize the model before the integration continues. The EnKF originated as a version of the Extended Kalman Filter (EKF) ([43],[44]) for large problems. The classical KF [36] method is optimal in the sense of minimizing the variance only for linear systems and Gaussian statistics. Similar to the particle filter method, the EnKF stems from a Monte Carlo integration of the Fokker-Planck equation governing the evolution of the PDF that describes the prior, forecast, and error statistics. In the analysis step, each ensemble member is updated according to the KF scheme and replaces the covariance matrix by the sample covariance computed from the ensemble. However, the EnKF presents two potential problems namely: 1) Even though the EnKF uses full non-linear dynamics to propagate the forecast error statistics, the EnKF assumes that all probability distributions involved are Gaussian. 2) The updated ensemble preserves only the first two moments of the posterior. Let $p(\mathbf{x})$ denote the Gaussian prior probability density distribution of the state vector \mathbf{x} with mean μ and covariance \mathcal{Q}

$$p(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^T \mathcal{Q}^{-1}(\mathbf{x} - \mu)\right)$$

We assume the data \mathbf{z} to have a Gaussian PDF with covariance \mathcal{R} and mean $\mathcal{H}\mathbf{x}$, where \mathcal{H} is the so-called the observation matrix, is related to h of equation (5), and where the value $\mathcal{H}\mathbf{x}$ assumes the value of the data \mathbf{z} would be for the state \mathbf{x} in absence of observation errors. Then the conditional probability or likelihood $p(\mathbf{z}|\mathbf{x})$ assumes the form

$$p(\mathbf{z}|\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{z} - \mathcal{H}\mathbf{x})^T \mathcal{R}^{-1}(\mathbf{z} - \mathcal{H}\mathbf{x})\right).$$

According to the Bayes theorem the posterior probability density follows from the relation

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})p(\mathbf{x}). \quad (16)$$

There are many variants of implementing the EnKF of various computational efficiency and in what follow we employ standard formulation of the EnKF for linear and nonlinear observation operators with covariance localization. See ([37], [38], [42] [41], [55]). The implementation of the standard EnKF may be divided into three steps, as follows:

- **Setting and matching**

- Define the ensemble

$$\mathcal{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \quad (17)$$

be an $n_x \times N$ matrix whose columns are a sample from the prior distribution. N being the number of the ensemble members.

- Form the ensemble mean

$$\bar{\mathcal{X}} = \mathcal{X} \cdot \mathbf{1}_N, \quad (18)$$

where $\mathbf{1}_N \in \mathbb{R}^{N \times N}$ is the matrix where each element is equal to 1.

- Define the ensemble perturbation matrix \mathcal{X}' and set the $\mathbb{R}^{n_x \times n_x}$ ensemble covariance matrix \mathcal{C}

$$\mathcal{X}' = \mathcal{X} - \frac{1}{N} \bar{\mathcal{X}}, \quad (19)$$

$$\mathcal{C} = \frac{\mathcal{X}' \mathcal{X}'^T}{N - 1}, \quad (20)$$

- **Sampling**

- Generate

$$\mathcal{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N] \quad (21)$$

be an $n_z \times N$ matrix whose columns are a replicate of the measurement vector \mathbf{z} plus a random vector from the normal distribution $\mathcal{N}(0, \mathcal{R})$.

- Form the $\mathbb{R}^{n_z \times n_z}$ measurement error covariance

$$\mathcal{R} = \frac{\mathcal{Z} \mathcal{Z}^t}{N - 1}, \quad (22)$$

- **Updating** Obtain the posterior \mathcal{X}^p by the linear combinations of members of the prior ensemble

$$\mathcal{X}^p = \mathcal{X} + \mathcal{C} \mathcal{H}^T (\mathcal{H} \mathcal{C} \mathcal{H}^T + \mathcal{R})^{-1} (\mathcal{Z} - \mathcal{H} \mathcal{X}) \quad (23)$$

The matrix

$$\mathcal{K} = \mathcal{C} \mathcal{H}^T (\mathcal{H} \mathcal{C} \mathcal{H}^T + \mathcal{R})^{-1} \quad (24)$$

is the Kalman gain matrix. Since \mathcal{R} is always positive definite(i.e. covariance matrix), the inverse $(\mathcal{H} \mathcal{C} \mathcal{H}^T + \mathcal{R})^{-1}$ exists. An easy computation shows that the mean and covariance of the posterior or updated ensemble are given by

$$\bar{\mathcal{X}}^p = \mathcal{X}^p + \mathcal{K} [\mathbf{z} - (\mathcal{H} \mathcal{X}^p + \mathbf{d})], \quad (25)$$

and

$$\mathcal{C}^p = \mathcal{C} - \mathcal{K} [\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R}] \mathcal{K}^T, \quad (26)$$

the vector \mathbf{d} which appears in (25) stems from the affine measurement relation

$$h(\mathbf{x}) = \mathcal{H}\mathbf{x} + \mathbf{d}. \quad (27)$$

In the case of nonlinear observation operators, a modification to the above algorithm is advised. As presented in Evensen [39], let $\hat{\mathbf{x}}$ the augmented state vector made of the state vector and the predicted observation vector (nonlinear in this case).

$$\hat{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathcal{H}(\mathbf{x}) \end{pmatrix}. \quad (28)$$

Define the linear observation operator $\hat{\mathcal{H}}$ by

$$\hat{\mathcal{H}} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{y} \quad (29)$$

and carry out the steps of the EnKF formulation in augmented state space $\hat{\mathbf{x}}$ and $\hat{\mathcal{H}}$ instead of \mathbf{x} and \mathcal{H} . Superficially, this technique appears to reduce the nonlinear problem to the previous linear observation operator case. However, whilst the augmented problem, involving linear observation problem, is a reasonable way of formulating the EnKF, it is not as well-founded as the linear case, which can be justified as an approximation to the exact and optimal KF.

To prevent the occurrence of filter divergence usually due to the background-error covariance estimates from small number of ensemble members as pointed out in Houtekamer and Mitchell [47], the use of covariance localization was suggested. Mathematically, the covariance localization increases the effective rank of the background error covariances. See the work of Gaspari and Cohn [50] also Hamill and Snyder [48], Hamill [49] and Ehrendorfer [53]. The covariance localization consists of multiplying point by point the covariance estimate from the ensemble with a correlation function that is 1.0 at the observation location and zero beyond some prescribed distance. Mathematically, to apply covariance localization, the Kalman gain

$$\mathcal{K} = \mathcal{C}\mathcal{H}^T (\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1}$$

is replaced by a modified gain

$$\hat{\mathcal{K}} = [\rho \circ \mathcal{C}] \mathcal{H}^T (\mathcal{H} [\rho \circ \mathcal{C}] \mathcal{H}^T + \mathcal{R})^{-1} \quad (30)$$

where $\rho \circ$ denotes the Schur product (The Schur product of matrices \mathcal{A} and \mathcal{B} is a matrix \mathcal{D} of the same dimension, where $d_{ij} = a_{ij}b_{ij}$) of a matrix \mathcal{S} with local support with the covariance model generated by the ensemble. Various correlation matrices have been used, for horizontal localization Gaspari and Cohn [50] constructed a Gaussian-shaped function that is actually a fourth-order piece-wise polynomial. Houtekamer and Mitchell [47] and Evensen [39] used a cut-off radius so that observations are not assimilated beyond a certain distance from the grid points.

3.5. The Maximum Likelihood Ensemble Filter

The Maximum Likelihood Ensemble Filter (MLEF) proposed by Zupanski [66], and Zupanski and Zupanski [67] is a hybrid filter combining the 4-D variational method with the EnKF.

It maximizes the likelihood of posterior probability distribution which justifies its name. The method comprises three steps, a forecast step that is concerned with the evolution of the forecast error covariances, an analysis step based on solving a non-linear cost function and updating step.

- **Forecasting:** It consists of evolving the square root analysis error covariance matrix through the ensembles. The starting point is from the evolution equation of the discrete Kalman filter described in Jazwinski [43]

$$P_f^k = \mathcal{M}_{k-1,k} P_a^{k-1} \mathcal{M}_{k-1,k}^T + Q_{k-1}, \quad (31)$$

where P_f^k is the forecast error covariance matrix at time k , $\mathcal{M}_{k-1,k}$ the non-linear model evolution operator from time $k-1$ to time k , and Q_{k-1} is the model error matrix which is assumed to be normally distributed. Since P_a^{k-1} is positive matrix for any k , equation (31) could be factorized and written as

$$P_f^k = \overbrace{\left(\mathcal{M}_{k-1,k} (P_a^{k-1})^{1/2} \right)}^{(P_f^{k-1})^{1/2}} \left(\mathcal{M}_{k-1,k} (P_a^{k-1})^{1/2} \right)^T + Q_{k-1} \quad (32)$$

where $(P_a^{k-1})^{1/2}$ is of the form

$$(P_a^{k-1})^{1/2} = \begin{pmatrix} p_{(1,1)}^{k-1} & p_{(2,1)}^{k-1} & \cdots & p_{(N,1)}^{k-1} \\ p_{(1,2)}^{k-1} & p_{(2,2)}^{k-1} & \cdots & p_{(N,2)}^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{(1,n)}^{k-1} & p_{(2,n)}^{k-1} & \cdots & p_{(N,n)}^{k-1} \end{pmatrix}, \quad (33)$$

as usual N is the number of ensemble members and n the number of state variables. Using equation(33), the square root forecast error covariance matrix $(P_f^{k-1})^{1/2}$ can then be expressed as

$$(P_f^{k-1})^{1/2} = \begin{pmatrix} b_{(1,1)}^{k-1} & b_{(2,1)}^{k-1} & \cdots & b_{(N,1)}^{k-1} \\ b_{(1,2)}^{k-1} & b_{(2,2)}^{k-1} & \cdots & b_{(N,2)}^{k-1} \\ \vdots & \vdots & \ddots & \vdots \\ b_{(1,n)}^{k-1} & b_{(2,n)}^{k-1} & \cdots & b_{(N,n)}^{k-1} \end{pmatrix}, \quad (34)$$

where for each $1 \leq i \leq N$

$$\mathbf{b}_i^{k-1} = \begin{pmatrix} b_{(i,1)}^{k-1} \\ b_{(i,2)}^{k-1} \\ \vdots \\ b_{(i,n)}^{k-1} \end{pmatrix} = \mathcal{M}_{k-1,k} \begin{pmatrix} x_1^{k-1} + p_{(i,1)}^{k-1} \\ x_2^{k-1} + p_{(i,2)}^{k-1} \\ \vdots \\ x_n^{k-1} + p_{(i,n)}^{k-1} \end{pmatrix} - \mathcal{M}_{k-1,k} \begin{pmatrix} x_1^{k-1} \\ x_2^{k-1} \\ \vdots \\ x_n^{k-1} \end{pmatrix}. \quad (35)$$

The vector $\mathbf{x}^{k-1} = \begin{pmatrix} x_1^{k-1} \\ x_2^{k-1} \\ \vdots \\ x_n^{k-1} \end{pmatrix}$ is the analysis state from the previous assimilation cycle. which is found from the posterior analysis pdf as presented in Lorenc [45].

- **Analyzing:** The analysis step for the MLEF involves solving a non-linear minimization problem. As in Lorenc [45], the associated cost function is defined in terms of the forecast error covariance matrix and is given as

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T (P_f^k)^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2} [\mathbf{y} - h(\mathbf{x})]^T \mathcal{R}^{-1} [\mathbf{y} - h(\mathbf{x})] \quad (36)$$

where \mathbf{y} is the vector of observations, h is the non-linear observation operator, \mathcal{R} is the observational covariance matrix and \mathbf{x}_b is a background state given by

$$\mathbf{x}_b = \mathcal{M}_{k-1,k}(\mathbf{x}^{k-1}) + Q_{k-1}. \quad (37)$$

Through a Hessian preconditioner we introduce the change of variable

$$(\mathbf{x} - \mathbf{x}_b) = (P_f^{k-1})^{1/2} (\mathcal{I} + \mathbb{H}\text{ess})^{-T/2} \xi \quad (38)$$

where ξ is vector of control variables, $\mathbb{H}\text{ess}$ is the Hessian matrix of \mathcal{J} and \mathcal{I} is the identity matrix. The Hessian matrix $\mathbb{H}\text{ess}$ is provided by

$$\mathbb{H}\text{ess} = (P_f^{k-1})^{T/2} \tilde{\mathcal{H}}^T \tilde{\mathcal{H}} \mathcal{R}^{-1} \tilde{\mathcal{H}} (P_f^{k-1})^{T/2} = (\mathcal{R}^{-1/2} \tilde{\mathcal{H}} (P_f^{k-1})^{1/2})^T (\mathcal{R}^{-1/2} \tilde{\mathcal{H}} (P_f^{k-1})^{1/2}) \quad (39)$$

here $\tilde{\mathcal{H}}$ is the Jacobian matrix of the non-linear observation operator h evaluated at the background state \mathbf{x}_b .

To overcome the difficulty of calculating $\mathbb{H}\text{ess}$, which is principally due to the non-linear observation operator, (P_f^{k-1}) is used.

Let \mathcal{Z} the matrix defined by

$$\mathcal{Z} = \begin{pmatrix} z_{(1,1)} & z_{(2,1)} & \cdots & z_{(N,1)} \\ z_{(1,2)} & z_{(2,2)} & \cdots & z_{(N,2)} \\ \vdots & \vdots & \ddots & \vdots \\ z_{(1,n)} & z_{(2,n)} & \cdots & z_{(N,n)} \end{pmatrix}, \quad \mathbf{z}_i = \begin{pmatrix} z_{(i,1)} \\ z_{(i,2)} \\ \vdots \\ z_{(i,n)} \end{pmatrix} = \mathcal{R}^{-1/2} \tilde{\mathcal{H}} \mathbf{b}_i^{k-1}, \quad (40)$$

using the approximation

$$\mathbf{z}_i \approx \mathcal{R}^{-1/2} [h(\mathbf{x} + \mathbf{b}_i^{k-1}) - h(\mathbf{x})], \quad (41)$$

the matrix $\mathbb{H}\text{ess}$ can then be approximated by

$$\mathbb{H}\text{ess} = \mathcal{Z} \mathcal{Z}^T. \quad (42)$$

The approximation (42) is not necessary and a derivation of the MLEF not involving (42) has been recently developed in Zupanski et al. [68]

- **Updating:** The final point about MLEF is to update the square root analysis error covariance matrix

$$(P_a^k)^{T/2} = (P_f^{k-1})^{T/2} (\mathcal{I} + \mathbb{H}\text{ess}(\mathbf{x}_{opt}))^{-T/2}, \quad (43)$$

where \mathbf{x}_{opt} is approximately the minimum of the cost function \mathcal{J} given by (36).

4. Numerical Experiments

The novelty of this work consists in comparing for the first time the above described and commonly used filters in the framework of the K-S equation model focusing on performance in the presence of nonlinear observation operators.

4.1. Numerical Solution of the K-S Equation

We consider the one-dimensional PDE with initial data as used in [20] and [21]

$$\begin{cases} u_t = -u_{xxxx} - u_{xx} - uu_x, & x \in [0, 32\pi] \\ u(x + L, t) = u(x, t), & L = 32\pi, \forall t > 0 \\ u(x, 0) = \cos\left(\frac{x}{16}\right) \left(1 + \sin\left(\frac{x}{16}\right)\right) \end{cases} \quad (44)$$

The system (44) is known to be stiff. In fact, the stiffness is due to rapid exponential decay of some modes (the dissipative part), the stiffness is also due to rapid oscillations of some modes (the dispersive part).

As the equation is periodic, a Fourier spectral method is used for spatial discretization. Despite the remarkable success of the spectral and pseudo-spectral methods for a wide range of applications [22] and [23], the set of ODEs for the mode amplitudes is stiff, due to the time scale associated with the n th mode scales as $O(n^{-m})$ for large n , where m is the order of the highest spatial derivative, so that the highest modes evolve on short time scales.

In order to carry out numerical solution of K-S, a modification of the exponential time-differencing fourth-order Runge-Kutta method (ETDRK4) has been used. This method has been proposed by Cox and Matthews [25] and further modified by Kassam and Trefethen [20]. A short review of the ETDRK4 is as follows:

First we transform (44) to Fourier space

$$\hat{u}_t = -\frac{ik}{2}\hat{u}^2 + (k^2 - k^4)\hat{u} \quad \text{with } \hat{u}^2 = \sum_{k' \in \mathbb{Z}} \hat{u}_{k'}\hat{u}_{k-k'}, \quad (45)$$

set

$$\mathcal{L}\hat{u}(k) = (k^2 - k^4)\hat{u}(k), \quad \mathcal{N}(\hat{u}, t) = \mathcal{N}(\hat{u}) = -\frac{ik}{2}(\mathcal{F}(\mathcal{F}^{-1}(\hat{u}^2))), \quad (46)$$

\mathcal{L} and \mathcal{N} stand for linear and nonlinear operators, respectively. \mathcal{F} denotes the discrete Fourier transform. Write (45) in an operational form

$$\hat{u}_t = \mathcal{L}\hat{u} + \mathcal{N}(\hat{u}, t). \quad (47)$$

Define $v = e^{-\mathcal{L}t}u$ where $e^{-\mathcal{L}t}$ the integrating factor to obtain

$$v_t = e^{-\mathcal{L}t}\mathcal{N}(e^{\mathcal{L}t}v). \quad (48)$$

Let h denote the time step length, then integrating (48) we obtain

$$u_{n+1} = e^{\mathcal{L}h}u_n + e^{\mathcal{L}h} \int_0^h e^{-\mathcal{L}\tau} \mathcal{N}(u(t_n + \tau), t_n + \tau) d\tau, \quad (49)$$

where u_n is the solution at the time $t = nh$ and $0 < \tau < h$.

The equation (49) is exact, and the various order EDT schemes differ only on the way one approximates the integral in (49). Cox and Matthews [25] proposed the generating formula

$$u_{n+1} = e^{\mathcal{L}h}u_n + h \sum_{m=0}^{s-1} g_m \sum_{k=0}^m (-1)^k \binom{m}{k} \mathcal{N}_{n-k} \quad (50)$$

where s is the order of the scheme. The coefficients g_m are provided by the recurrence relation

$$\begin{cases} \mathcal{L}(hg_0) = e^{\mathcal{L}h} - I, \\ \mathcal{L}(hg_{m+1}) + \mathbf{I} = g_m + \frac{1}{2}g_{m-1} + \frac{1}{3}g_{m-2} + \cdots + \frac{1}{m+1}g_0, \quad m \geq 0. \end{cases} \quad (51)$$

We solve the K-S equation employing 64 Fourier spectral modes and integrate from $t = 0$ to $t = 250$ (nondimensional time units) using the EDTRK4 time stepping. The time evolution for the K-S equation is depicted in Fig.(1), while the time evolution for the K-S at two different locations is presented in Fig.(2).

4.2. Design of the Numerical experiments

The following table summarizes the logic of the numerical experiments conducted in this paper

Filter	Features	Observation operator type
PF	resampling and merging methods	linear and nonlinear
EnKF	standard and with localization and inflation	linear and nonlinear
MLEF	standard	linear and nonlinear

In order to compare the above mentioned EnKF, PF and MLEF sequential DA methods we use the K-S equation model with both linear and nonlinear observation operators for the same number of ensemble members. During the numerical experiments the following set-up for each of the above mentioned DA methods was used. The observations were provided at a frequency consisting of one observation every 30 time steps. Each run was carried out for 250 nondimensional time units with a time step of $\Delta t = 1/10$. Thus 80 DA cycles were used for each experiment. The number of Fourier modes in (2) was set to 64 modes. The standard deviation of each of the state noise and observation noise was taken to be $\sqrt{2}$.

For each type of the observation operator, the conclusion on the performance of each of the above mentioned three methods from the numerical experiments will be based on:

- amount of discrepancy between the time evolution of the forecast and the filter solutions
- reduction of the analysis covariance error
- reduction of the root mean squared error (RMSE) and the χ^2 statistics

4.3. Sequential data assimilation with a linear observation operator

Before comparing the performance above mentioned DA methods, we examine the performance of each method separately starting with the PF method in the presence of linear observation operator.

In Fig.(3) we present the time evolution of the weight distribution of particles (ensemble members) using the systematic resampling method using 300 particles. The results obtained yielded a similar behavior to that observed when multinomial or residual resampling methods were employed. We also provide in Fig.(4) a detailed representation of the weight distribution for a fixed time, here $t = T_{final} = 250$. The Fig.(4) shows clearly that despite resampling most of the particles are degenerated and only few have significant weights. It illustrates the fact discussed by Snyder et al. [63] that an exponential growth (i.e. $\exp(\tau^2/2)$, where τ^2 is

the variance of the observation log-likelihood) in the number of particles is required for the particle filter to perform satisfactorily.

Fig.(5) illustrates two different (modes) locations for the above mentioned PF experiment. The trends show an almost perfect matching between the forecast and the PF filter data assimilation results. This fact being further confirmed by the small value of the analysis error covariance depicted in Fig.(6). Moreover, Fig.(6) shows how the analysis error covariance decreases and becomes localized (concentration around the diagonal) with time. 1000 particles were employed and the plots were taken after $t = 20, t = 100$, and $t = 250$ time units respectively. We present in Fig.(7) the true and the PF solutions for different number of particles and for fixed time $t = 250$ time units. From Fig.(7) one can deduce that as the number of particles increases, the PF solution tends to converge towards the true solution.

Even though the purpose of the paper is not to compare the efficiency of resampling methods, we have examined the residual, systematic and the multinomial resampling methods. They all yield similar results. In terms of CPU time, the SR resampling method was the fastest followed by the RR resampling method. This is in good agreement with the finding of Arulampalam et al. ([32]) and also Doucet et al. ([31]).

We also examined and compared the performance the Merging Particle Filter (MPF) of Nakano et al. [57] to the standard PF with resampling. Indeed, the MPF method was faster than the PF method, but for the K-S equation model the PF provides more accurate results as it can be easily seen from the Fig.(8).

An identical set-up of the time step and the observation operator has been employed for the standard EnKF sequential data assimilation experiment. The number of ensemble members employed in EnKF has been set to 100.

We present in Fig.(9) a perfect matching between the forecast and the EnKF solution at two different locations. The analysis error covariance for different times is depicted in Fig. (10). Similar to the PF case, the analysis error covariance decreases and becomes local as time increases. The effect on the number of ensemble members is illustrated in Fig.(11). Indeed, as the number of ensemble members increases, the EnKF data assimilation solution converges towards the true solution. It should be also pointed out that there exists a certain number of ensemble members, say N_{em} , such that for $\forall n > N_{em}$ the discrepancy between the EnKF solution and the true solution is virtually the same. This claim was observed to be valid throughout our numerical tests.

The effect of localization and inflation in EnKF for the case of the K-S equation model is presented in Fig.(12), however this effect has no sizable impact for our model in comparison to the numerical weather prediction (NWP) models employed by Hamill et al [48], Houtekamer and Mitchell [47] and Jardak et al. [64]. The additive covariance inflation of Anderson and Anderson [46] with $r = 1.001$ has been employed. The covariance localization was also used to prevent filter divergence. For the present model, we tested the fifth order correlation function of Gaspari and Cohn [50] and the usual Gaussian correlation function

$$\rho(D) = \exp\left[-\left[\frac{D}{l}\right]^2\right], l = 450 \text{ length units} \quad (52)$$

which was found to yield better results.

Our numerical results using the MLEF are presented in Fig.(13) and Fig.(14). Again, the analysis error covariance decreases and becomes localized (concentration around the diagonal) with time as shown in Fig.(14).

Table I. Root mean square error for different number of particle/ensemble members, linear observation operator case

	PF	EnKF	MLEF
70	1.3558	2.0961	1.38189133
100	1.4313	0.3297	1.43104518
150	1.3635	1.4603	1.7722971
250	1.3903	1.3388	1.5577693

Table II. χ^2 statistics for different number of particle/ensemble members, linear observation operator case

	PF	EnKF	MLEF
70	2.5346690	2.345658	1.9612679
100	1.9854678	1.136783	2.0121631
150	3.0457823	1.783459	2.6414165
250	1.934876	1.655631	1.8208084

Comparing linear observation operator results of covariances, rmse and χ^2 for the three above mentioned sequential data assimilation filters leads to the conclusion that for the model at hand, the EnKF provides better results followed by the MLEF and then the PF filter with SR resampling. These results are of course a function of our particular implementation of the EnKF, PF and MLEF and are function of the number of particles (number of ensemble members) and the number of observations as well as the particular model tested. In as far as computational efficiency the EnKF was found to be faster than the MLEF, while the PF filter proved to be the most time consuming in as far as the K-S equation model with linear observation operator is concerned. The performance of the above motioned methods is assessed and encapsulated in tables Table(I) and Table(II).

4.4. Sequential data assimilation with nonlinear observation operator

A similar setting to the linear observation case in terms of number of ensemble member and particles has been used to carry out numerical experiments with nonlinear observation operator. The EnKF version tailored to nonlinear observation operator as described in section 3 has been implemented and an observation operator assuming the form $\mathcal{H}(u) = u^2$ has been considered. As depicted in Fig.(15) an intense discrepancy between the forecast and EnKF mean solutions is easily seen. We join many authors, Nakano et al. [57] among them to draw the conclusion that even with the modification suggested by Evensen [39], the EnKF does not provide good estimates in the case of nonlinear observation operator applied to the K-S equation model.

Having the previous conclusion in mind, the comparison reduces to comparing the PF and the MLEF methods in the presence of nonlinear observation operator. Both the PF and the MLEF exhibited good performances in terms of sequential data assimilation as measured in different error metrics such as error covariance matrix, rms error and χ^2 statistics. The PF method with SR resampling appears to have an edge in terms of CPU time, while in other

Table III. Root mean square error for different number of particle/ensemble members, nonlinear observation operator case

	PF for $h(u) = u^2$	MLEF for $h(u) = u^2$	PF for $h(u) = exp$	MLEF for $h(u) = exp(u)$
70	0.35639178	1.3781648	0.27696167	1.43203425
100	0.3244502203	1.4070521	0.30416423	1.40807023
250	0.366861214	1.43104518	0.31519531	1.5476146

Table IV. χ^2 statistics for different number of particle/ensemble members, nonlinear observation operator case

	PF for $h(u) = u^2$	MLEF for $h(u) = u^2$	PF for $h(u) = exp$	MLEF for $h(u) = exp(u)$
70	2.6344987	3.7372720	4.3238997	10.1312923
100	2.8783455	4.2008870	3.7356629	9.37517203
250	1.9873456	2.0121631	2,1099875	9.52631038

metrics, the PF has a slight edge over the MLEF method as seen in Table(III), Table(IV) and the figures Fig.(16), Fig.(17), Fig.(18) and Fig.(19).

This result is further confirmed, as it can also be assessed from tables III, IV, for the case when the nonlinear observation operator assumes the form $h(u) = exp(u)$. It appears that the high CPU time required by the MLEF is due to the minimization component of this method. This remark remains to be elucidated in further research. The results of the performance of the PF and MLEF methods with $h(u) = exp(u)$ as nonlinear observation operator are presented in Table(III), Table(IV). The results of our numerical experiments are presented in Fig.(20), Fig.(21) and Fig.(22).

5. Summary & conclusions

In this work we have presented a comparison of 3 sequential data assimilation methods, PF, EnKF and MLEF applied to the K-S equation model. For each of the aforementioned methods both linear and nonlinear observation operators have been employed and extensive numerical experiments have been carried out to assess their relative performances.

It appears that the standard EnKF implementation with localization and inflation works well for the case of linear observation operator. However, it has a major drawback dealing with nonlinear observation operators. The PF and the MLEF perform satisfactory for both types of observation operators. Our experience suggests that the PF with resampling overall performed somewhat better than the MLEF for the particular set-up tested. Additional future research will focus on alternative PF resampling such as the one proposed in Xiong et al. [56] and Salman [51]. Also, new options of using hybrid PF EnKF filter as proposed by Brockmeier [52] will be considered to alleviate the degeneracy catastrophe of PF.

ACKNOWLEDGEMENTS

The research of Prof. I.M. Navon and Dr. M. Jardak is supported by the National Science Foundation (NSF), grant ATM-03727818. The authors also acknowledge the support by NASA Modeling, Analysis, and Prediction Program under Award NNG06GC67G. The first author is thankful to W.H.Hanya for the continuous help.

REFERENCES

1. Adams RA. *Sobolev Spaces* Academic Press, 1975.
2. Mazya VG. *Sobolev Spaces* Springer-Verlag, 1985.
3. Lions JL. *Quelques méthodes de résolution de problèmes aux limites non linéaires* Dunod: Paris, 1969.
4. Temam R. *Navier-Stokes Equations*, (3rd edn). North-Holland: Amsterdam New York, 1984.
5. Temam R. *Infinite-dimensional dynamical systems in mechanics and physics*, *Applied Mat.Sci.* **68** Springer-Verlag, 1997.
6. Kuramoto Y, Tsuzuki T. On the formation of dissipative structures in reaction-diffusion systems. *Prog. Theor. Phys.* 1975; **54**:687–699
7. Kuramoto Y, Tsuzuki T. Persistent propagation of concentration waves in dissipative media far from thermal equilibrium. *Prog. Theor. Phys.* 1976; **55**:356–369
8. Sivashinsky GI. Nonlinear analysis of hydrodynamic instability in laminar flames I. Derivation of basic equations. *Acta Astrn.* 1977; **4**:1177–1206
9. Sivashinsky GI. On irregular wavy flow of a liquid film down a vertical plane. *Progr. Theoret. Phys.* 1980; **63**:2112–2114
10. Benny DJ. Long waves in liquid film. *J. Math and Phys.* 1966; **45**:150–155
11. Babchin AJ., Frenkel AL, Levich BG, Sivashinsky GI. Nonlinear saturation of Rayleigh-Taylor instability in thin films. *Phys. Fluids.* 1983; **26**:3159–3161
12. Hyman JM, Nicolaenko B. The Kuramoto-Sivashinsky Equation: A bridge between PDE'S and dynamical systems. *Physica D* 1986; **18**:113–126
13. Kevrekidis IG, Nicolaenko B, Scovel JC. Back in the saddle again: A computer assisted study of the Kuramoto-Sivashinsky equation. *SIAM J. Appl. Math.* 1990; Vol.50 **50**(3); 760–790
14. Foias C Nicolaenko B, Sell GR, Temam R. Inertial manifolds for the Kuramoto-Sivashinsky equation and an estimate of their lowest dimension. *J.Math.Pures Appl.* 1988; **67**:197–226
15. Nicolaenko B, Scheurer B, Temam R. Some global dynamical properties of the Kuramoto-Sivashinsky equations: Nonlinear stability and attractors. *Physica D.* 1985; **16**:155–185
16. Nicolaenko B, Scheurer B, Temam R. Some global dynamical properties of a class of pattern formation equations. *Comm. Partial Differential Equations* 1989; **14**: 245–297
17. Desertion KA, Kazantzis N. Improved performance of the controlled Kuramoto-Sivashinsky equation via actuator and controller switching. *Proceeding of the 2004 American Control Conference* 2004; 267–272
18. Lee CH, Tran HT. Reduced-order-based feedback control of the Kuramoto-Sivashinsky equation. *Journal of Computational and Applied Mathematics* 2005; **173**:1–19
19. Protas B, Bewley TR, Hagen G. A computational framework for the regularization of adjoint analysis in multiscale PDE systems, *J. Comput. Phys.* 2004; **195**(1):49–89
20. Kassam AK, Trefethen LN. Fourth-order time-stepping for stiff PDS's. *SIAM J.Sci.Compt.* 2005; **26**(4): 1214–1233
21. Krogstad S. Generalized integrating factor methods for stiff PDE's. *J.Comput.Phys.* 2005; **203**: 72–88
22. Canuto C, Hussaini MY, Quarteroni A, Zang TA. *Spectral methods in fluids dynamics* Springer-Verlag, 1988
23. Trefethen NL. *Spectral methods in Matlab*, Software. Environments. Tools , SIAM, Philadelphia, 2000
24. Boyd J P. Eight definitions of the slow manifold: Seiches, pseudoseiches and exponential smallness. *Dyn. Atmos. Oceans* 1995; **22**(1-2): 49–75
25. Cox SM, Matthews PC. Exponential time differencing for stiff systems. *J. Comput. Phys.* 2002; **176**: 430–455
26. Chorin AJ, Krause P. Dimensional reduction for a Bayesian filter. *PNAS* 2004; **101**no.42: 15013–15017
27. Hu C, Temam R. Robust boundary control for the Kuramoto-Sivashinsky equation. *Dynamics of Continuous Discrete and Impulsive Systems-Series B-Applications & Algorithms.* 2001; **8**(3): 315–338
28. Berliner ML, Wikle CK. A Bayesian tutorial for data assimilation. *Physica D* 2007; **30**:1–16
29. Berliner ML, Wikle C.K., Approximate importance sampling Monte Carlo for data assimilation. *Physica D* 2007; **30**:37–49
30. Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 2000; **10**(3):197–208

31. Doucet A, de Freitas JFG, Gordon NJ. *An introduction to sequential Monte Carlo methods*, in Sequential Monte Carlo Methods in Practice, Doucet A, de Freitas JFG, Gordon NJ (eds) Springer-Verlag: New York 2001.
32. Arulampalam MS, Maskell S, Gordon N, Clapp T. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking, *IEEE transactions on signal processing* 2002; **150**no.2:174–188
33. Gordon NJ, Salmund DJ, Smith AFM. Novel approach to nonlinear non-Gaussian Bayesian state estimate, *IEEE Proc. F* 1993; **140**:107–113
34. Kitagawa G. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics* 1996; **5**(1):1–25
35. Liu JS, Chen R. Sequential Monte Carlo Methods for Dynamic Systems *Journal of the American Statistical Association* 1998; **93**(442):1032–1044
36. Kalman RE. A new approach to linear filtering and prediction problems *Transaction of the ASME - Journal of basic Engineering* 1960; **Series D**(82):35–45
37. Evensen G. Sequential data assimilation in non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics *J. Geophys. Res.* 1994; **99**(C5 10):108–129
38. Burgers G, van Leeuwen PJ, Evensen G. Analysis scheme in the ensemble Kalman Filter *Mon. Weather Rev.* 1998; **126**:1719–1724
39. Evensen G. The Ensemble Kalman Filter: theoretical formulation and practical implementation *Ocean Dynamics* 2003; **53**:343–367
40. Evensen G. *Data Assimilation: The ensemble Kalman filter* Springer: Berlin, 2007.
41. Mandel J. A brief Tutorial on the ensemble Kalman filter manuscript, 2007.
42. Mandel J. Efficient Implementation of the Ensemble Kalman Filter *CCM Report* 2006; **231**
43. Jazwinski A H. *Stochastic processes and filtering theory* Academic Press: New York, 1970.
44. Bucy RS. Recurrent sets *Ann. Math. Statist.* 1965; **36**: 535–545.
45. Lorenc AC. "Analysis methods for numerical weather prediction," *Q. J. Roy. Meteor. Soc.* 1986; **112**: 1177–1194
46. Anderson JL, Anderson SL. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts *Mon. Wea. Rev.* 1999; **127**:2741–2785.
47. Houtekamer P, Mitchell M. Data assimilation using an ensemble Kalman filter technique *Mon. Wea. Rev.* 1998; **126**: 796–811
48. Hamill TM, Snyder C. A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme," *Mon. Wea. Rev.* 2000; **128**: 2905–2919
49. Hamill TM. Ensemble-based atmospheric data assimilation *Predictability of Weather and Climate*, Cambridge Press: 2006; 124–156
50. Gaspari G, Cohn SE. Construction of correlation functions in two and three dimensions *Quart. J. Roy. Meteor. Soc.* 1999; **125**: 723–757
51. Salman H. A hybrid grid/particle filter for Lagrangian data assimilation: Formulating the passive scalar approximation *Quart. J. Roy. Meteor. Soc.* 2008; **134**: 1539–1550
52. Brockmeier M. Using the particle filter for reservoir model updating, *Master of Science in Applied Mathematics, Delft University of Technology*; 2007.
53. Ehrendorfer E. A review of issues in ensemble-based Kalman filtering *Meteorologische Zeitschrift* 2007; **16**: 795–818
54. Hol JD, Schön TB, Gustafsson F. On Resampling Algorithms for Particle Filters *Nonlinear Statistical Signal Processing Workshop* 2006; Cambridge:United Kingdom
55. Lewis JL, Lakshivarahan S, Dhall S K. *Dynamic Data Assimilation: A least squares approach* in the series on Encyclopedia of Mathematics and its Applications **Volume 104** Cambridge University Press:2006
56. Xiong X, Navon IM, Uzunoglu B. A Note on the Particle Filter with Posterior Gaussian Resampling *Tellus* 2006; **58A**:456–460.
57. Nakano S, Ueno G, Higuchi T. Merging particle filter for sequential data assimilation *Nonlin. Processes Geophys.* 2007; **14**:395–409.
58. Kotecha G, Djurić PM, Gaussian particle filtering *IEEE Trans. Signal Processing* 2003; **51**:2592–2601.
59. Goldberg DE. *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley:Reading, 1989.
60. Silverman BW. *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, 1986.
61. Bengtsson T, Bickel P, Li B. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems *Probability and Statistics: Essays in Honor of David A. Freedman* 2008; **2**: 316–334
62. Bickel P, Li B, Bengtsson, T. Sharp Failure Rates for the Bootstrap Particle Filter in High Dimensions *IMS Collections Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh* 2008; **3**: 318–329
63. Snyder C, Bengtsson T, Bickel P, Anderson JL. Obstacles to high-dimensional particle filtering *Mon. Wea. Rev.* 2008; accepted

64. Jardak M, Navon IM, Zupanski D. Ensemble Data Assimilation for the Shallow water equation *J. Geophys. Res.* 2008; to be submitted
65. Zupanski D. A general weak constraint applicable to operational 4DVAR data assimilation systems *Mon. Wea. Rev.* 1997; **125**: 2274–2292
66. Zupanski M. Maximum Likelihood Ensemble Filter. Part1: the theoretical aspect *Mon. Wea. Rev.* 2005; **133**: 1710–1726
67. Zupanski D, Zupanski M. Model error estimation employing ensemble data assimilation approach *Mon. Wea. Rev.* 2006; **134**: 137–1354
68. Zupanski M, Navon IM, Zupanski D. The Maximum Likelihood Ensemble Filter as a non-differentiable minimization algorithm *Q.J.R. Meteorol.* 2008; **134**: 1039–1050

- Fig.1: Contours for the time evolution for the K-S equation
- Fig.2: K-S solution at different spatial positions
- Fig.3: Time evolution of the weights distribution. 300 particles over 250 time units, using the systematic resampling method, linear observation operator case
- Fig.4: Weights distribution at a fixed time. 300 particles at the final time $t = 250$, using the systematic resampling method, linear observation operator case
- Fig.5: Forecast and PF filter mean K-S solution at two different locations, linear observation operator case
- Fig.6: Analysis error covariance using 200 particles. (a) after $t = 20$ time units , (b) after $t = 100$ time units and (c) after $t = 250$ time units, linear observation operator case
- Fig.7: PF solutions for different number of particles (ensemble members) taken at $t = 250$ time units, linear observation operator case
- Fig.8: PF with the SR resampling and MPF solutions for the K-S equation model, linear observation operator case
- Fig.9: Forecast and EnKF filter mean K-S solution at two different locations, linear observation operator case
- Fig.10: Analysis error covariance using 100 ensemble members. (a) after $t = 20$ time units , (b) after $t = 100$ time units and (c) after $t = 250$ time units, linear observation operator case
- Fig.11: EnKF solutions for different number of ensemble members, linear observation operator case
- Fig.12 : EnKF solutions with and without the inflation and localization covariance feature, linear observation operator case
- Fig.13: Contours of the analysis error covariance after 250 time units using 100 ensembles members, linear observation operator case
- Fig.14: Analysis error covariance using 100 ensemble members. (a) after $t = 20$ time units , (b) after $t = 100$ time units and (c) after $t = 250$ time units, linear observation operator case
- Fig.15: Means of the forecast and the EnKF filtered K-S solution in the case of a nonlinear observation operator $\mathcal{H}u = u^2$, nonlinear observation operator case
- Fig.16: Means of the K-S solution using SIR particle filter, nonlinear observation operator case
- Fig.17: Analysis error covariance using 100 particles. (a) after $t = 20$ time units , (b) after $t = 100$ time units and (c) after $t = 250$ time units, nonlinear observation operator case
- Fig.18: Contours of the analysis error covariance after 250 time units using 100 ensemble members, nonlinear observation operator case
- Fig.19: Analysis error covariance using 100 ensemble members. (a) after $t = 20$ time units , (b) after $t = 100$ time units and (c) after $t = 250$ time units, nonlinear observation operator case
- Fig.20: Means of the K-S solution using particle filter, nonlinear observation operator case
- Fig.21: Contours of the analysis error covariance after 250 time units using 100 ensemble members, nonlinear observation operator case
- Fig.22: Analysis error covariance using 100 ensemble members. (a) after $t = 20$ time units , (b) after $t = 100$ time units and (c) after $t = 250$ time units, nonlinear observation operator case

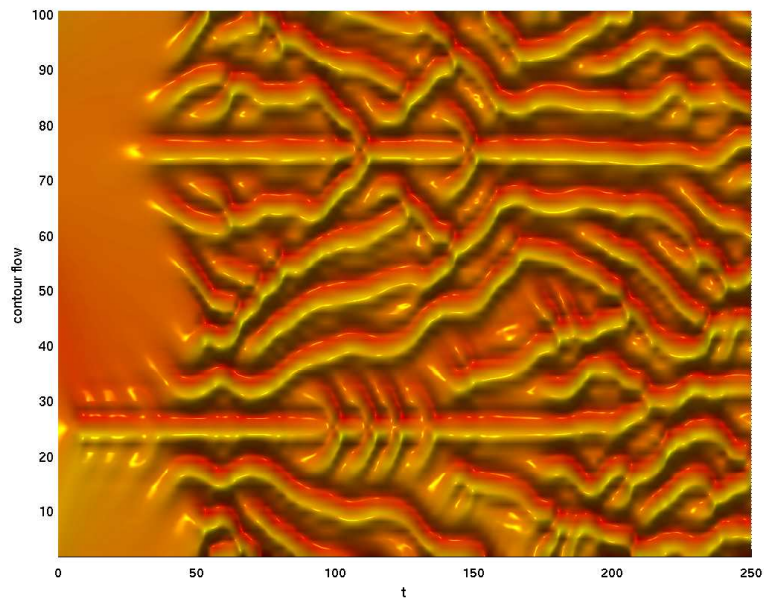


Figure 1. Contours for the time evolution for the K-S equation

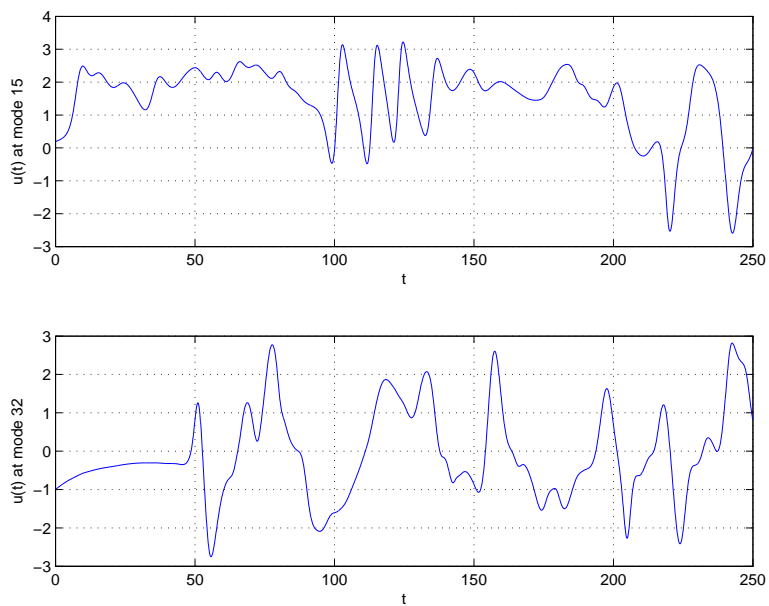


Figure 2. K-S solution at different spatial positions

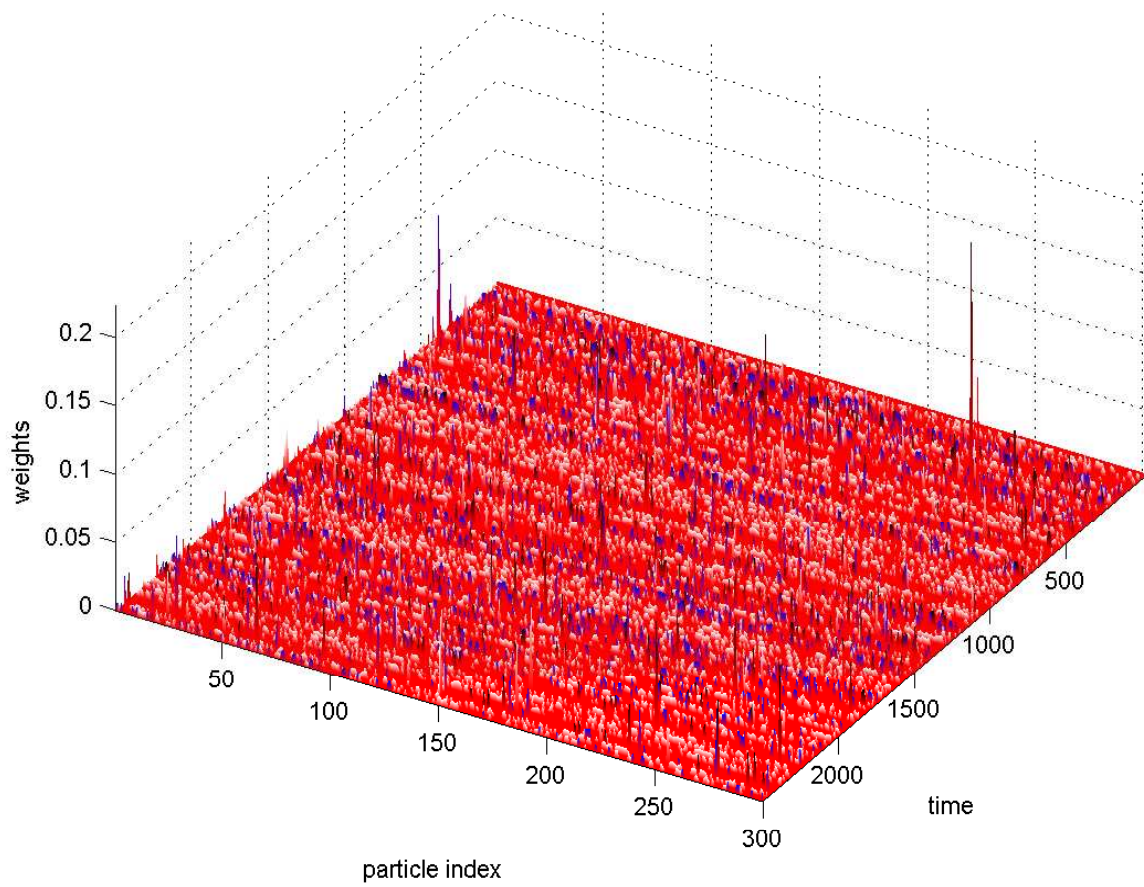


Figure 3. Time evolution of the weights distribution. 300 particles over 250 time units, using the systematic resampling method, linear observation operator case

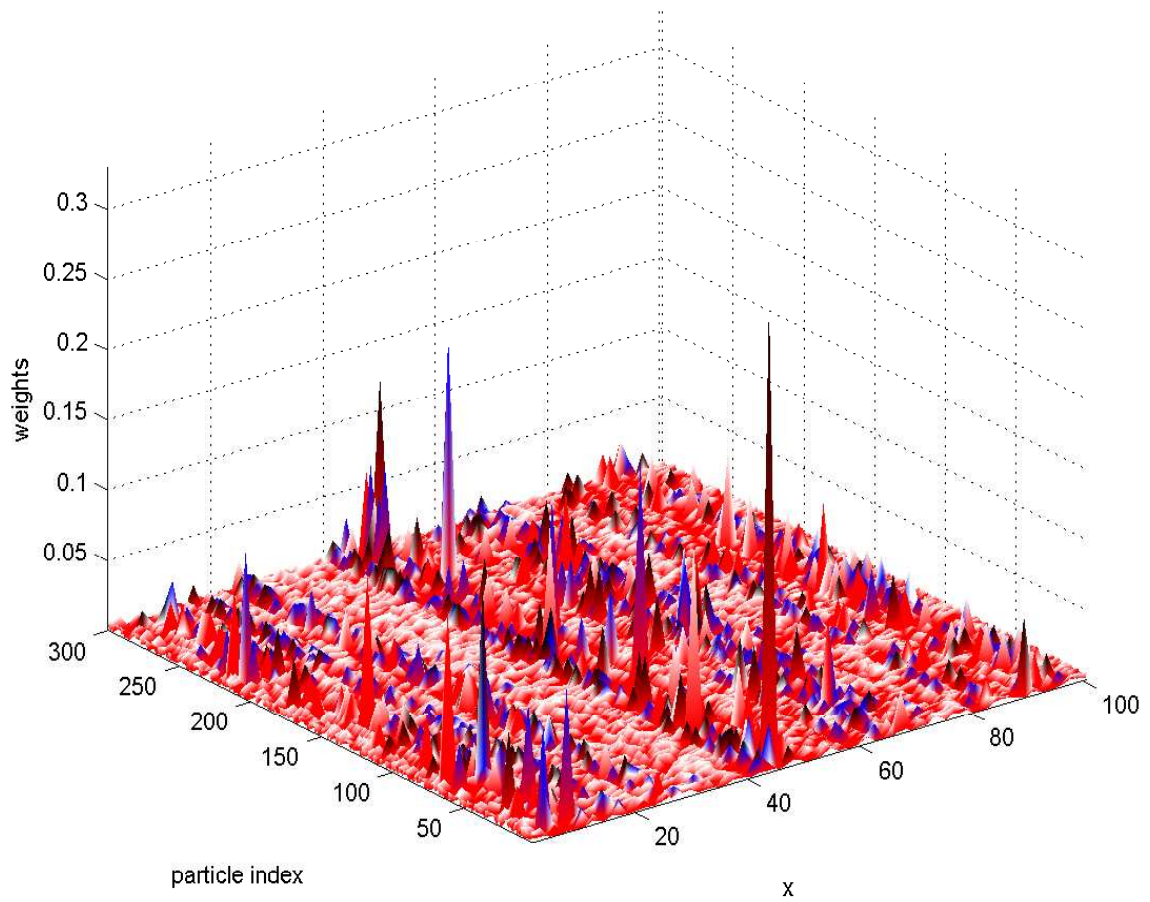


Figure 4. Weights distribution at a fixed time. 300 particles at the final time $t = 250$, using the systematic resampling method, linear observation operator case

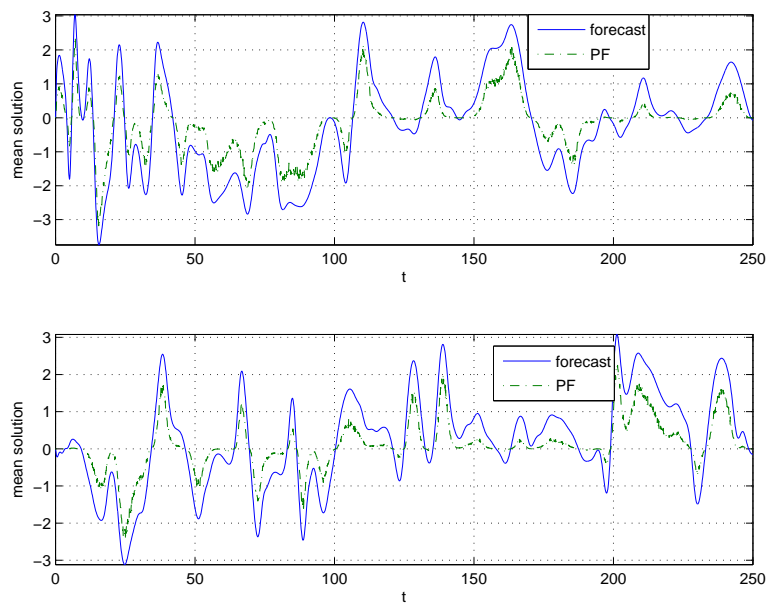


Figure 5. Forecast and PF filter mean K-S solution at two different locations, linear observation operator case

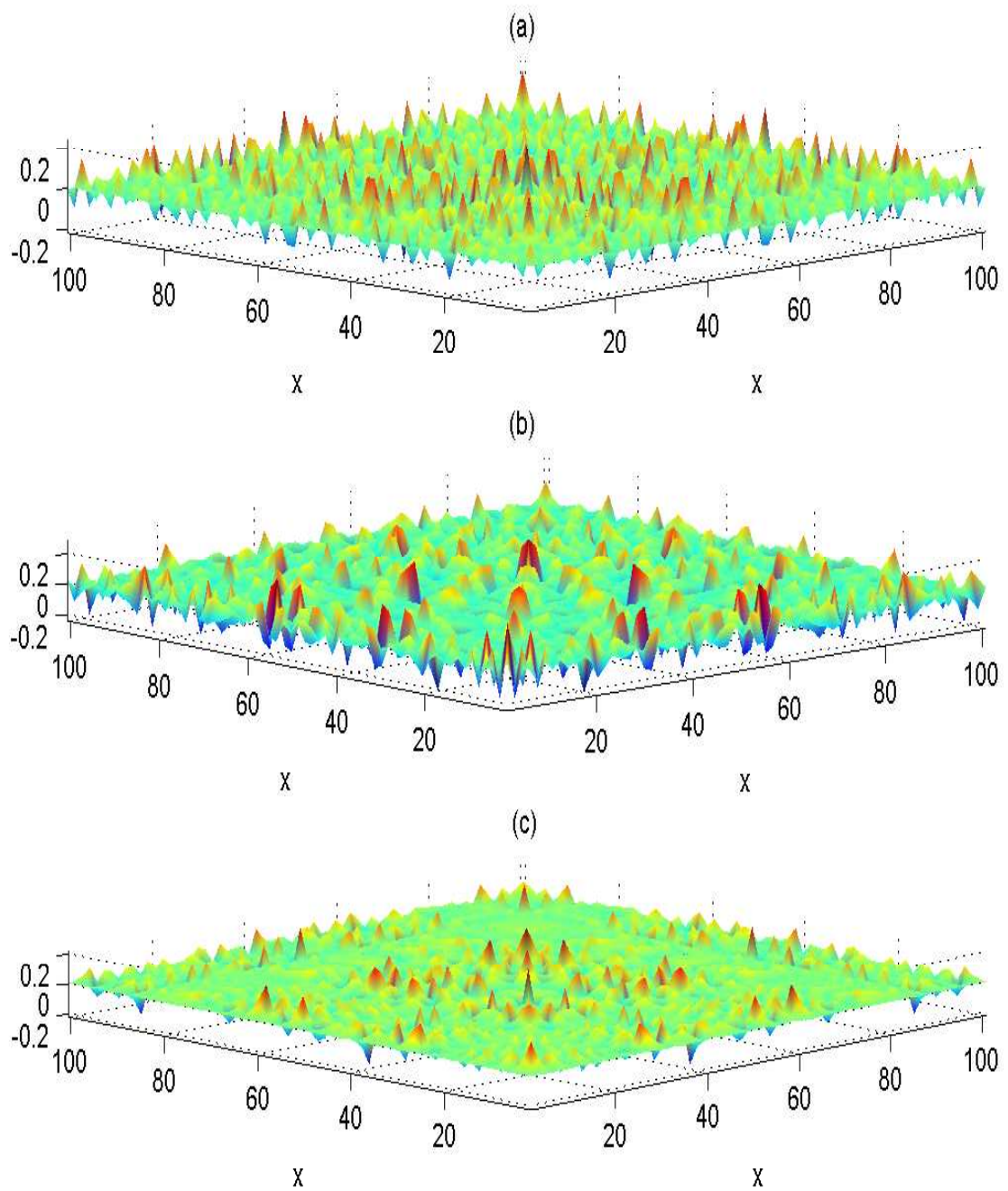


Figure 6. Analysis error covariance using 200 particles. (a) after $t = 20$ time units , (b) after $t = 100$ time units and (c) after $t = 250$ time units, linear observation operator case

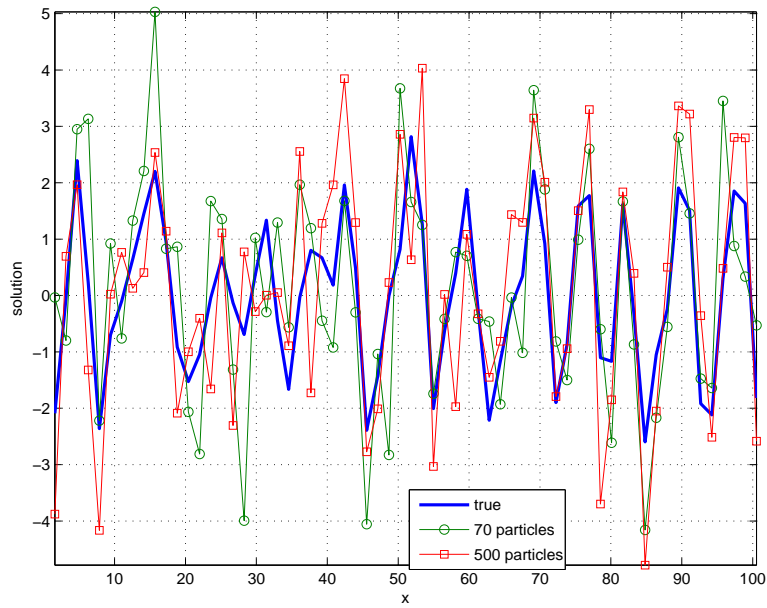


Figure 7. PF solutions for different number of particles (ensemble members) taken at $t = 250$ time units, linear observation operator case

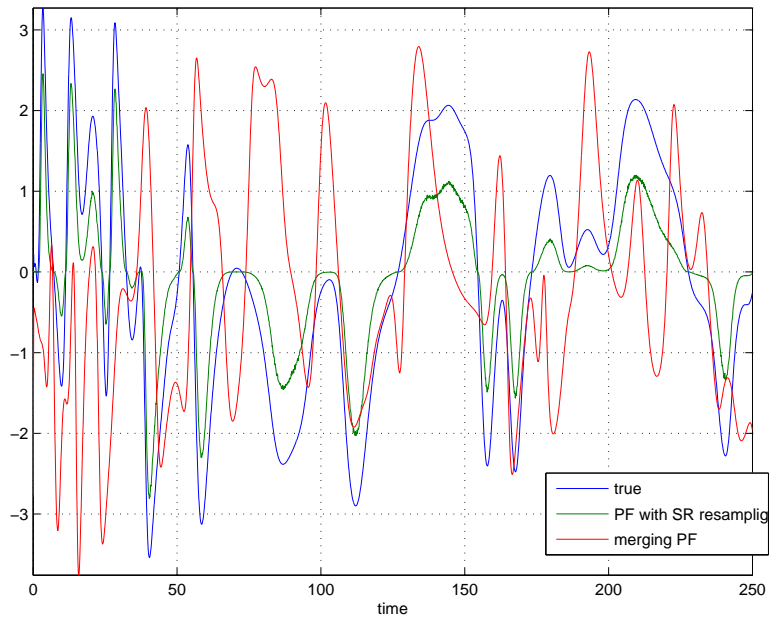


Figure 8. PF with the SR resampling and MPF solutions for the K-S equation model, linear observation operator case

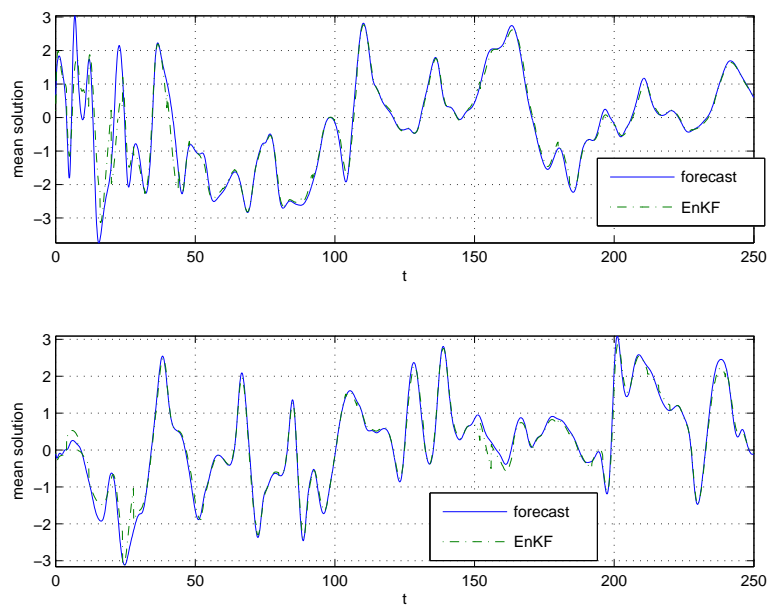


Figure 9. Forecast and EnKF filter mean K-S solution at two different locations, linear observation operator case

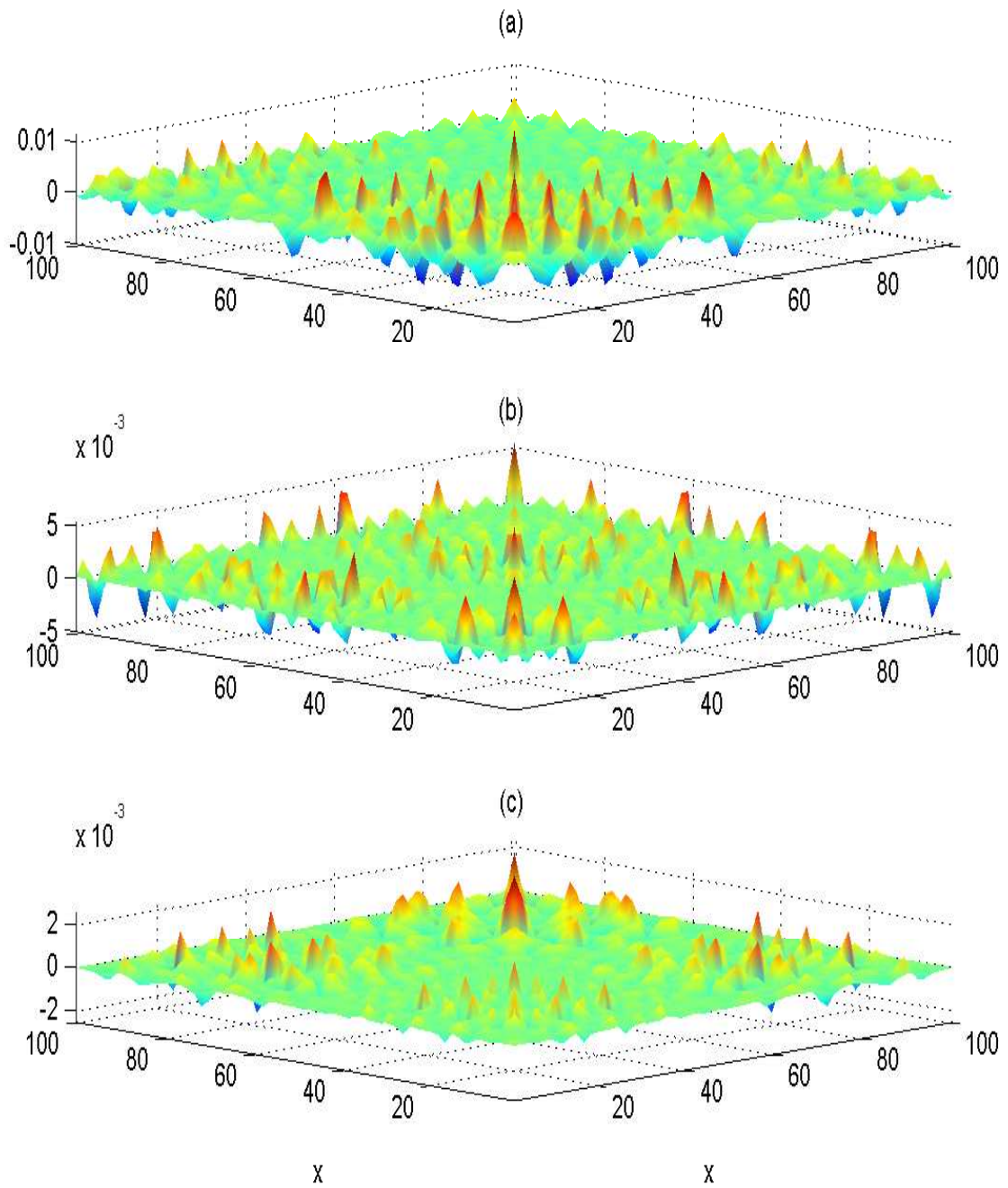


Figure 10. Analysis error covariance using 100 ensemble members. (a) after $t = 20$ time units , (b) after $t = 100$ time units and (c) after $t = 250$ time units, linear observation operator case

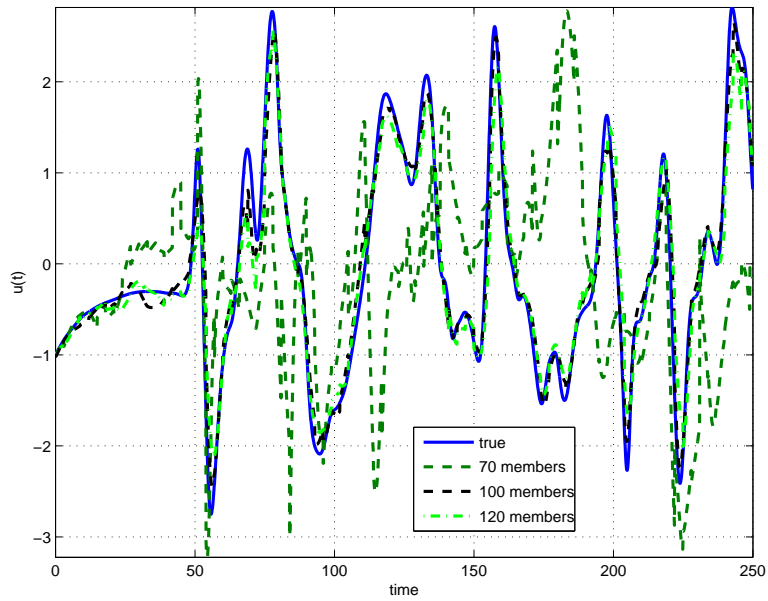


Figure 11. EnKF solutions for different number of ensemble members, linear observation operator case

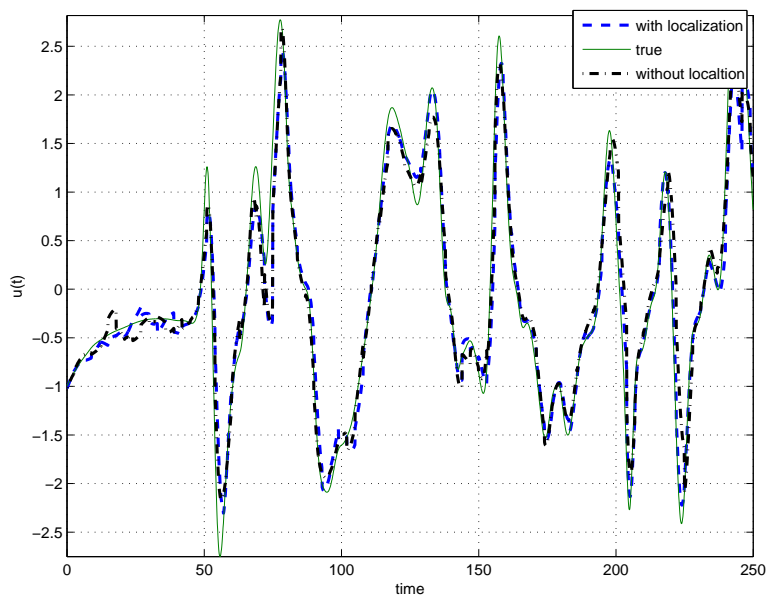


Figure 12. EnKF solutions with and without the inflation and localization covariance feature, linear observation operator case

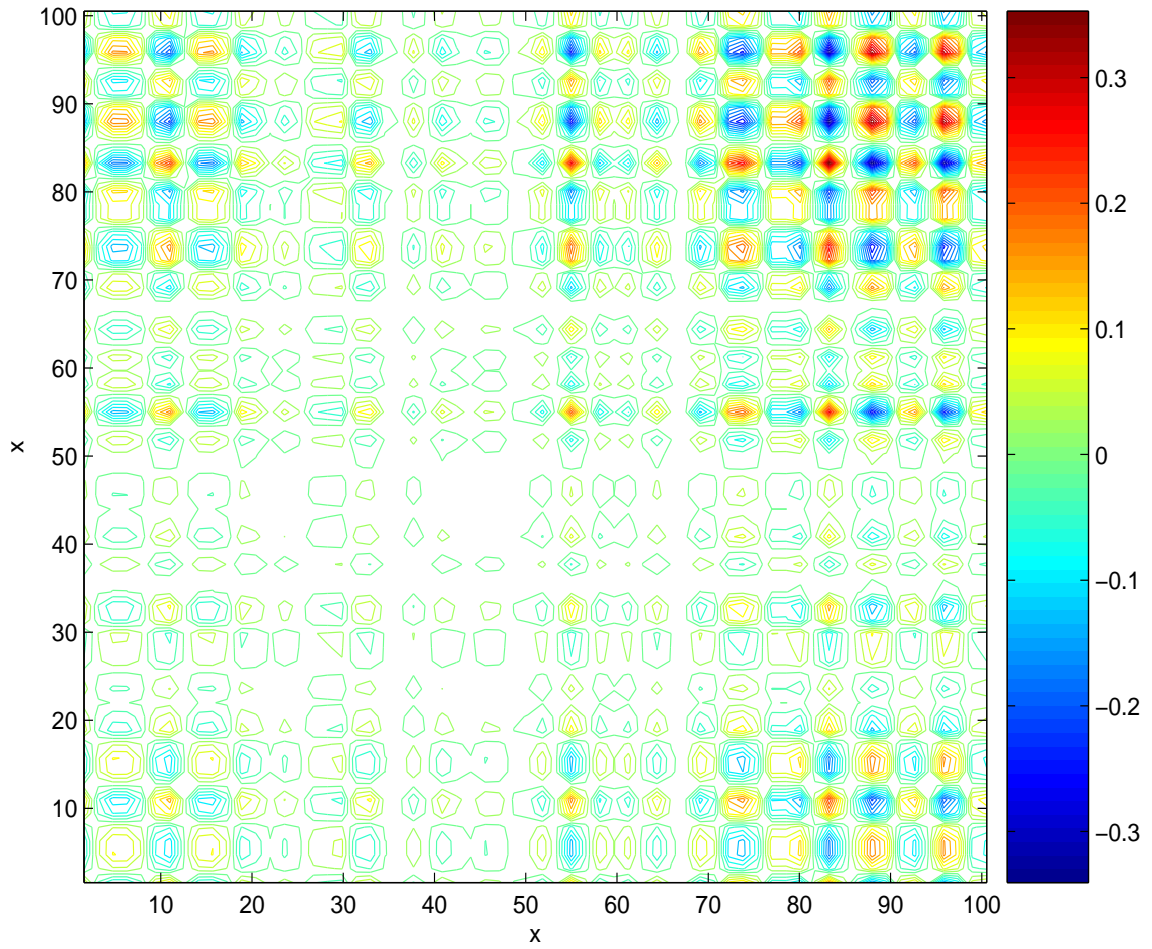


Figure 13. Contours of the analysis error covariance after 250 time units using 100 ensemble members, linear observation operator case

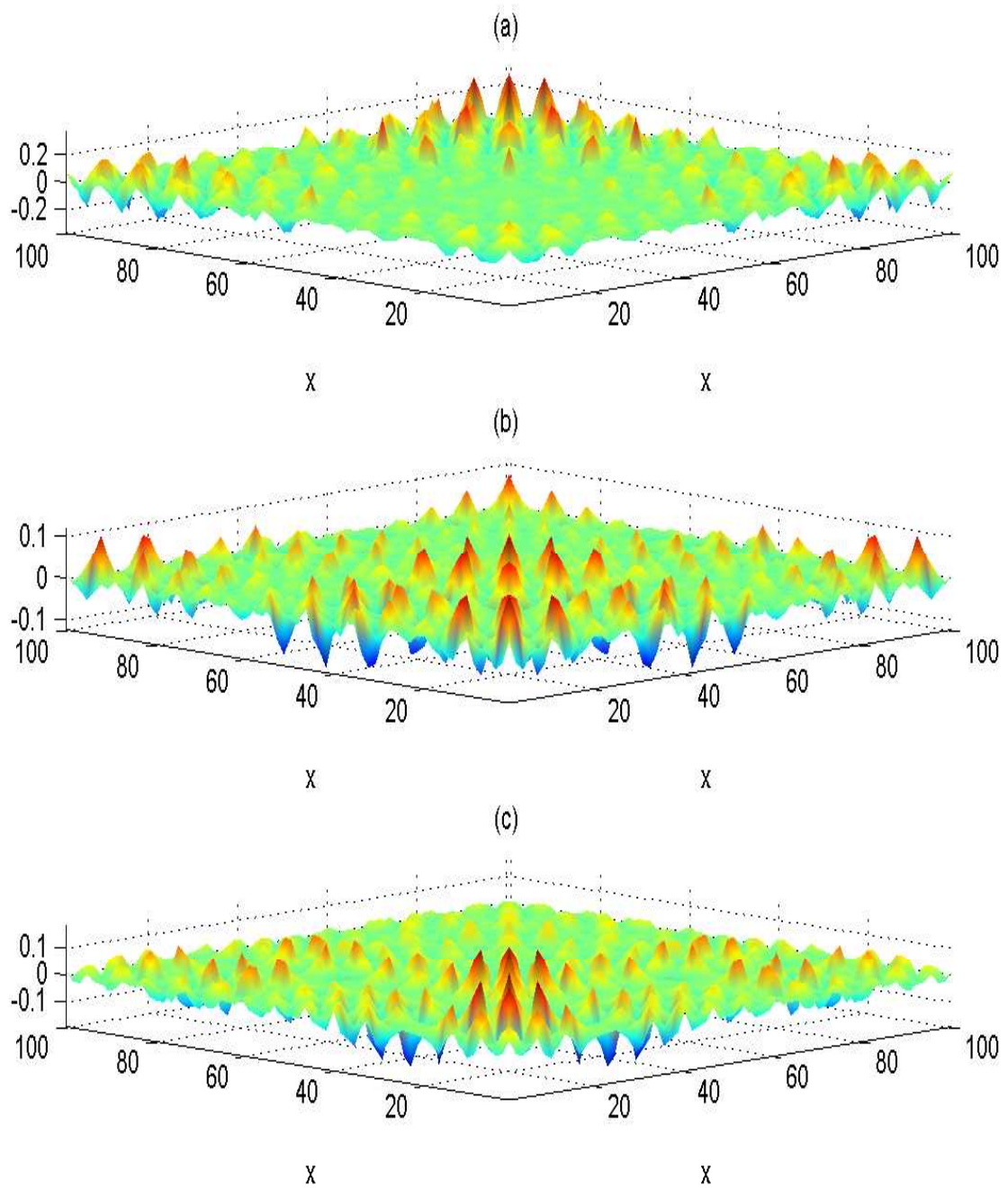


Figure 14. Analysis error covariance using 100 ensemble members. (a) after $t = 20$ time units, (b) after $t = 100$ time units and (c) after $t = 250$ time units, linear observation operator case

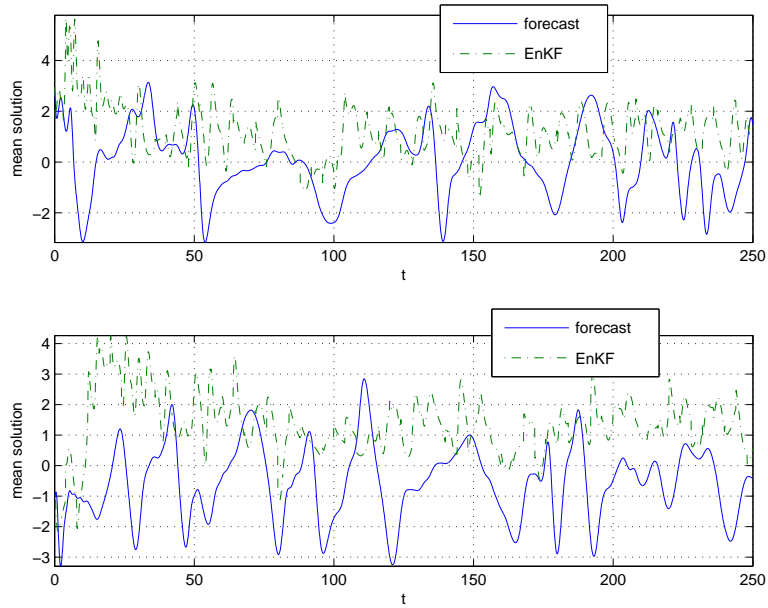


Figure 15. Means of the forecast and the EnKF filtered K-S solution in the case of a nonlinear observation operator $\mathcal{H}u = u^2$, nonlinear observation operator case

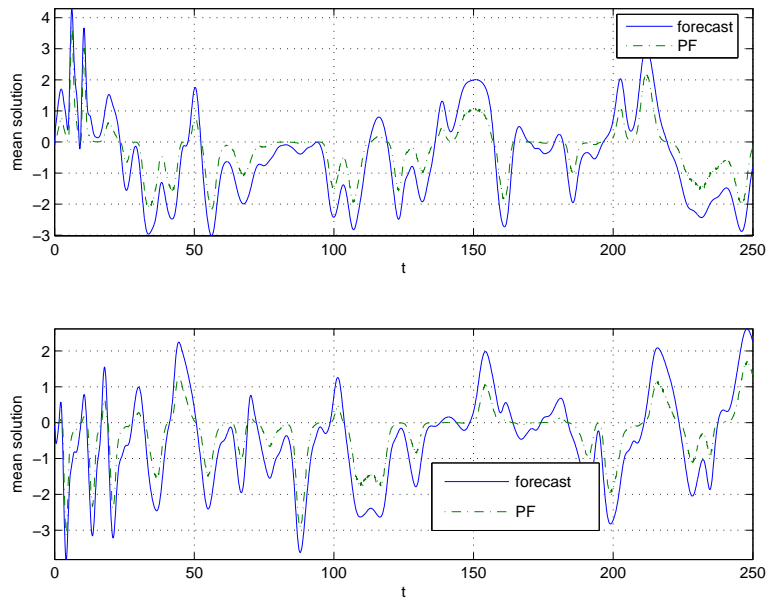


Figure 16. Means of the K-S solution using SIR particle filter, nonlinear observation operator case

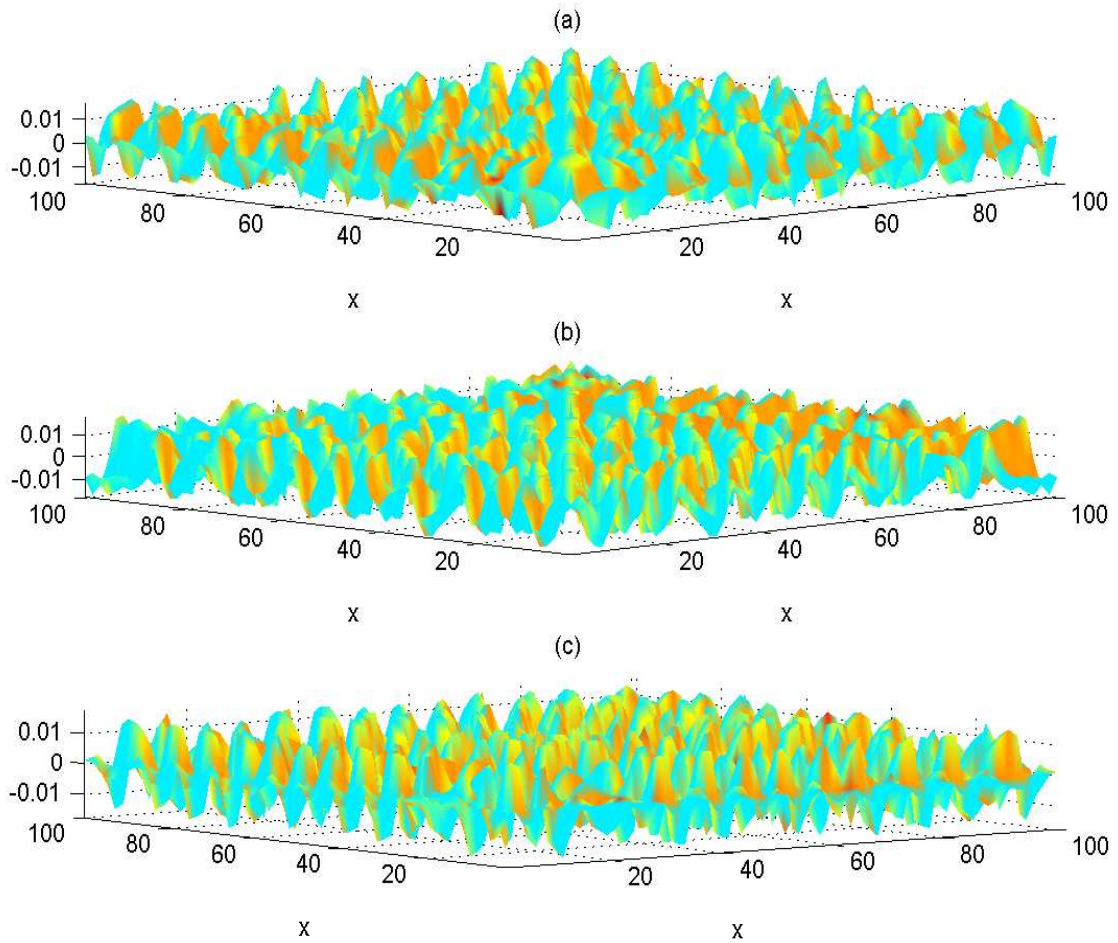


Figure 17. Analysis error covariance using 100 particles. (a) after $t = 20$ time units, (b) after $t = 100$ time units and (c) after $t = 250$ time units, nonlinear observation operator case

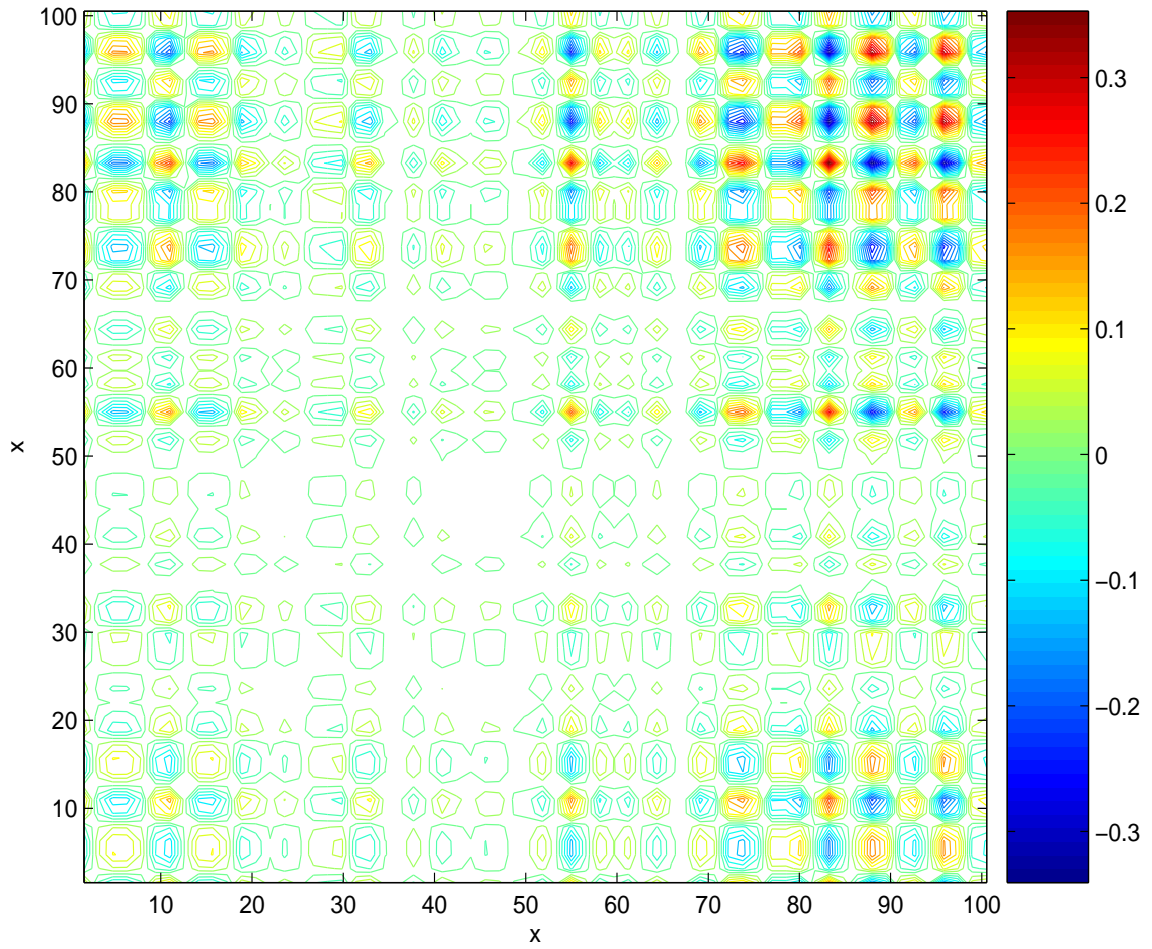


Figure 18. Contours of the analysis error covariance after 250 time units using 100 ensemble members, nonlinear observation operator case

